



Bollettino Notiziario - A.A. 2024/2025

LAUREA MAGISTRALE IN DATA SCIENCE (ORD. 2023)

Curriculum: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

ADVANCED DATABASES

Titolare: da definire

Periodo: I anno, 1 semestre

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: 24A+24E+12L; 5,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

BIG DATA MANAGEMENT

Titolare: da definire

Periodo: I anno, 2 semestre

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: 30A+15L; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

BIG DATA SEMINAR**Titolare:** da definire**Periodo:** I anno, 2 semestre**Indirizzo formativo:** BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics**Tipologie didattiche:** 21A; 2,00**Prerequisiti:**

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

BIOLOGICAL DATA**Titolare:** Prof. DAMIANO PIOVESAN**Mutuato da:** Laurea magistrale in Data Science (Ord. 2023)**Periodo:** II anno, 1 semestre**Indirizzo formativo:** BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics**Tipologie didattiche:** 48A; 6,00**Prerequisiti:**

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

BUSINESS PROCESS MANAGEMENT

Titolare: da definire

Periodo: I anno, 1 semestre

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: 24A+24E+12L; 5,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

DATA MINING

Titolare: da definire

Periodo: I anno, 1 semestre

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: 24A+24E+12L; 5,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

DATA WAREHOUSES

Titolare: da definire

Periodo: I anno, 1 semestre

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: 24A+24E+12L; 5,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

DATABASES SYSTEMS ARCHITECTURE

Titolare: da definire

Periodo: I anno, 1 semestre

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: 24A+12E+24L; 5,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

DEEP LEARNING AND HUMAN DATA ANALYTICS

Titolare: Prof. MICHELE ROSSI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: II anno, 1 semestre

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: 48A; 6,00

FINAL EXAMINATION

Titolare: da definire

Periodo: Il anno, annuale

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: 15A; 15,00

HUMANITIES: DEBATES ON ETHICS OF BIG DATA

Titolare: da definire

Periodo: I anno, 2 semestre

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: 24A; 2,00

HUMANITIES: FOREIGN LANGUAGE

Titolare: da definire

Periodo: I anno, 1 semestre

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: ; 5,00

HUMANITIES: FOREIGN LANGUAGE

Titolare: da definire

Periodo: I anno, 2 semestre

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: ; 2,00

INTERNSHIP

Titolare: Prof. TOMASO ERSEGHE

Periodo: Il anno, annuale

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: 15A; 15,00

Contenuti:

Durante il corso di laurea magistrale in Data Science, gli studenti svolgono uno stage presso un'azienda, un'industria, un centro di ricerca o un laboratorio universitario. Il tirocinio permette di sviluppare nuove competenze, entrare in contatto con l'ambiente di lavoro quotidiano, sviluppare la consapevolezza delle esigenze del mercato e dei clienti. Queste abilità sono molto apprezzate dal mercato del lavoro, rafforzano il curriculum degli studenti e migliorano le loro opportunità di lavoro. Siamo collegati con una serie di selezionati partner industriali e di ricerca, in Italia e nel mondo. Il tirocinio si svolge solitamente durante il 2° anno, 2° semestre e, seppur non obbligatorio, è spesso abbinato alla tesi di laurea magistrale. Una breve guida agli stage è disponibile qui: https://datascience.math.unipd.it/files/a-quick-guide-to-internships-@-data-science_final.pdf Una guida dettagliata per gli studenti iscritti è disponibile nella piattaforma elearning: <https://elearning.unipd.it/math/course/view.php?id=659> Per ulteriori richieste contattaci a stage.datascience@math.unipd.it

Testi di riferimento:

CONTENUTO NON PRESENTE

LAW AND DATA

Titolare: Dott.ssa FIORELLA DAL MONTE

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, 1 semestre

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

MACHINE LEARNING

Titolare: da definire

Periodo: I anno, 2 semestre

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: 36A+18L; 6,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

PROCESS MINING

Titolare: da definire

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, 1 semestre

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

SEMANTIC DATA MANAGEMENT**Titolare:** da definire**Periodo:** I anno, 2 semestre**Indirizzo formativo:** BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics**Tipologie didattiche:** 30A+15L; 6,00**Prerequisiti:**

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

STATISTICAL LEARNING**Titolare:** Prof. ALBERTO ROVERATO**Mutuato da:** Laurea magistrale in Data Science (Ord. 2023)**Periodo:** II anno, 1 semestre**Indirizzo formativo:** BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics**Tipologie didattiche:** 48A; 6,00**Prerequisiti:**

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

STOCHASTIC METHODS

Titolare: Prof. MARCO FERRANTE

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, 1 semestre

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

TIME-SERIES ANALYSIS FOR BUSINESS ECONOMIC AND FINANCIAL DATA

Titolare: Prof.ssa MARIANGELA GUIDOLIN

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, 1 semestre

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

VIABILITY OF BUSINESS PROJECTS

Titolare: da definire

Periodo: I anno, 2 semestre

Indirizzo formativo: BDMA - Big Data Management and Analytics - Statistics & Deep Learning for Data Analytics

Tipologie didattiche: 36A+18L; 6,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

Curriculum: Corsi comuni

Curriculum: Data Science - Piano di studio Biological Data Analytics

ADVANCED TOPICS IN DATA SCIENCE

Titolare: Prof. LAMBERTO BALLAN

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, annuale

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 24A; 3,00

BIG DATA COMPUTING

Titolare: Prof. FRANCESCO SILVESTRI

Mutuato da: Scuola Galileiana di Studi Superiori - Classe di Scienze Naturali

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

BIOLOGICAL DATA

Titolare: Prof. DAMIANO PIOVESAN

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Conoscenze base di informatica, algoritmi di ottimizzazione e machine learning. Linguaggio di programmazione: Python.

Conoscenze e abilità da acquisire:

Il corso intende comunicare le conoscenze di base sui diversi tipi di dati biologici, come sequenze, strutture, reti e letteratura. Inoltre intende permettere allo studente lo svolgimento autonomo di un progetto di ricerca in bioinformatica, definendo lo stato dell'arte per un problema aperto e un tentativo di risolverlo con lo sviluppo di software che estenda librerie esistenti e la valutazione critica dei risultati ottenuti.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso si compone di lezioni frontali, esercitazioni pratiche al computer, sviluppo di un progetto e presentazione dello stesso con discussione critica. Le esercitazioni servono per familiarizzare lo studente con le librerie software da usare per un progetto bioinformatico relativo ad un problema attuale diverso per ogni gruppo. La presentazione del progetto richiede una discussione in cui far emergere i punti di forza e debolezza della soluzione prescelta.

Contenuti:

Il corso si compone di diverse parti, che corrispondono ai diversi tipi di dati biologici: Sequenze - DNA e proteine - Allineamenti - Ricerca veloce in database Strutture - Proteine non globulari - Motifs discovery - Predizione di disordine Funzione - Knowledgebases - Predizione di funzione Letteratura - Artticolli scientifici - Text mining

Modalità di esame:

L'esame è suddiviso in quattro parti separate, ognuna delle quali deve essere superata per completare con successo il corso. I pesi relativi di ciascuna parte sono indicati tra parentesi di seguito: 1. Test scritto a metà corso, che consiste in domande teoriche e pratiche (20%) 2. Journal Club, presentazione di un articolo scientifico inerente al corso (20%) 3. Progetto software (30%) 4. Presentazione del progetto e valutazione critica (30%) Per superare l'esame, gli studenti devono completare con successo tutte e quattro le parti e ottenere un voto sufficiente in ciascuna di esse. Il test scritto a metà corso valuterà la comprensione degli studenti dei concetti teorici affrontati nel corso, mentre il progetto software valuterà la loro capacità di applicare questi concetti a un problema pratico. La presentazione del progetto valuterà invece la capacità degli studenti di comunicare i risultati della loro ricerca e di valutare il lavoro dei loro colleghi.

Criteri di valutazione:

- Comprensione di concetti e gli algoritmi presentati a lezione - Capacità di applicare le nozioni fornite a lezione su problemi reali - Capacità critica di saper utilizzare i metodi nei modi più opportuni, scegliendo tra le alternative possibili - Capacità di sviluppare software riutilizzabile estendendo librerie esistenti - Capacità espositiva e di discussione critica

Testi di riferimento:

Krogh, Anders; Eddy, Sean R.; Mitchison, Graeme; Durbin, Richard; Krogh, Anders, Biological sequence analysis probabilistic models of proteins and nucleic acids. : , 1998

Eventuali indicazioni sui materiali di studio:

Sul sito E-learning vengono resi disponibili molti materiali per il corso. Questi comprendono i lucidi del corso (appena disponibili) e le registrazioni audio (podcast), le dispense e la letteratura usata per i progetti. Le lezioni pratiche sono organizzate tramite CoLab notebook in Python.

DEEP LEARNING (MATICOLE DISPARI)

Titolare: Prof. ALESSANDRO SPERDUTI

Mutuato da: Laurea magistrale in Computer Science (Ord. 2021)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

DEEP LEARNING (MATICOLE PARI)

Titolare: Prof. ALESSANDRO SPERDUTI

Mutuato da: Laurea magistrale in Computer Science (Ord. 2021)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

ENGLISH LANGUAGE B2 (PRODUCTIVE SKILLS)

Titolare: Prof. WOLFGANG ERB

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, annuale

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: ; 3,00

FINAL EXAMINATION

Titolare: da definire

Periodo: Il anno, annuale

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 15A; 15,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

FUNDAMENTALS OF INFORMATION SYSTEMS

Titolare: Prof. GIORGIO MARIA DI NUNZIO

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 96A; 12,00

HIGH DIMENSIONAL PROBABILITY FOR DATA SCIENCE

Titolare: Prof. ALBERTO CHIARINI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

INTERNSHIP

Titolare: da definire

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, annuale

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: ; 18,00

INTRODUCTION TO HUMAN DISEASES

Titolare: Prof. DENIS DOMINIQUE MARTINVALET

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Conoscenze base di statistica e programmazione

Conoscenze e abilità da acquisire:

Il corso fornisce un'introduzione alla fisiopatologia umana e ai principali meccanismi responsabili dello sviluppo di malattie rare, multifattoriali e complesse quali il cancro, malattie del neurosviluppo, malattie autoimmuni, reazioni di ipersensibilità e invecchiamento. A seconda delle dimensioni della classe, organizzeremo alcune attività pratiche di laboratorio per testare la capacità delle cellule immunitarie di uccidere le cellule tumorali bersaglio.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso prevede lezioni frontali, esercitazioni di casi di studio e lo sviluppo di un progetto di gruppo e presentazione dello stesso con discussione critica.

Contenuti:

Il corso è composto da due parti: 1. Introduzione alla fisiologia umana e alle malattie complesse UN. Introduzione alla biochimica e alla biologia cellulare B. Introduzione alla fisiologia umana C. Principi di ereditarietà e mutazioni patogene D. Meccanismi molecolari di patologie complesse: cancro, malattie del neurosviluppo, malattie autoimmuni, reazioni di ipersensibilità e invecchiamento

Modalità di esame:

L'esame riguarda la valutazione delle conoscenze biologiche sulla biologia umana e cellulare relative alle malattie complesse.

Criteri di valutazione:

criteri di valutazione sono: Comprensione dei concetti presentati a lezione La capacità di applicare i concetti descritti a problemi reali La capacità di presentazione e discussione critica

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

Text books Rubin's Pathology: Clinicopathologic Foundations of Medicine, Seventh Edition Robbins & Cotran: Pathologic Basis of Disease, Ninth Edition Peter Parham: The Immune System, Fourth Edition Exame Multiple choice & Open questions

INTRODUCTION TO MOLECULAR BIOLOGY

Titolare: Prof.ssa MARIA PENNUTO

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Nessuno

Conoscenze e abilità da acquisire:

Gli studenti apprenderanno come vengono generati i dati di "omics", da quali tipi di campioni biologici derivano (DNA, RNA, proteine), nozioni da integrare con la teoria dell'evoluzione di Darwin e le leggi della genetica di Mendel. Si introdurrà il concetto di mutazione, da distinguere dal concetto di Single Nucleotide Polymorphism (SNP). Agli studenti verranno forniti esempi di come si ottengono i dati di omics in lab. Impareranno come l'evoluzione e la pressione selettiva modificano la genetica degli esseri viventi e la trasmissione dell'informazione genetica nell'equilibrio dinamico che governa i processi biologici. Gli studenti apprenderanno il dogma della biologia, che vede la conservazione dell'informazione genetica nel DNA, molecola da cui viene prodotto l'RNA, da cui a sua volta vengono prodotte le proteine, e come l'informazione genetica è conservata nel nostro genoma e trasmessa alla progenie. Inoltre, gli studenti impareranno come e perchè l'espressione genica differisce da un tessuto ad un altro, facendo sì che lo stesso gene venga differenzialmente espresso in tessuti diversi. In questo corso gli studenti inizieranno dalla teoria dell'evoluzione fino ad arrivare al concetto di gene, mutazioni, genetica e leggi di Mendel. Gli studenti otterranno nozioni di biologia molecolare e biochimica al fine di capire da quali fonti -campioni biologici e materiale biologico (DNA, RNA, proteina)- si ottengono i dati di omics. Lo studente farà esperienze di laboratorio al fine di verificare come si manipolano DNA, RNA e proteine, come si clonano i geni (ingegneria genetica) e come questi si esprimono in cellula.

Attività di apprendimento previste e metodologie di insegnamento:

L'insegnante farà utilizzo di diapositive (file ppt) che saranno disponibili agli studenti sulla piattaforma moodle. Si effettueranno ripassi al fine di approfondire specifici argomenti. Si faranno verifiche ad intervalli al fine di valutare lo stato di apprendimento degli studenti. Per ciò che riguarda le esperienze di laboratorio, gli studenti effettueranno esperimenti di manipolazione degli acidi nucleici e uso di colture cellulari.

Contenuti:

Il corso ha lo scopo di fornire agli studenti le conoscenze e gli strumenti necessari a comprendere i seguenti aspetti: GENETICA: 1. Leggi di Mendel (genetica mendeliana) ed eccezioni alle leggi di Mendel (genetica non mendeliana): caratteri ereditari. 2. Cenni della teoria dell'evoluzione: da Lamarck a Darwin. 3. Cenni storici della scoperta del DNA: dalla trasmissione del carattere ereditario al concetto di GENE. 4. Il dogma della biologia molecolare: DNA->RNA->PROTEINA 5. La cellula: organelli, compartimenti, funzioni. DNA 1. Chimica: le basi azotate. 2. Struttura: la doppia elica. 3. Il codice genetico: il DNA è letto in triplette. 4. La replicazione del DNA. 5. Tecniche di purificazione del DNA: preparazioni di piccole e grandi quantità di DNA. 6. Tecniche di amplificazione del DNA: la PCR. 7. Tecniche di clonaggio dei geni: gli enzimi di restrizione e modificazione. RNA 1. Chimica: le basi azotate. 2. Struttura:

molecola a singolo filamento. 3. La trascrizione dell'RNA: regolazione dell'espressione genica. 4. Tecniche di purificazione dell'RNA e conversione in cDNA. 5. Tecniche di analisi dell'espressione genica su singolo gene e transcriptomics: la RT-PCR, microarrays, NGS. **PROTEINE** 1. Il codice genetico: universale e degenerato. 2. Il processo di traduzione: dall'RNA alla sintesi delle proteine. 3. Western blotting. **EPIGENETICA** 1. Organizzazione della cromatina 2. Codice istonico 3. Riprogrammazione cellulare: Dallo zigote all'organismo completo. 4. Riprogrammazione in lab: clonaggio cellulare. Il corso prevede esperienze di LABORATORIO con preparazione di campioni di DNA, RNA e proteine, espressione di geni target in cellula e analisi funzionale in cellule in coltura. Si prepareranno campioni per analisi delle proteine fornendo allo studente un esempio di come di prepara un campione per successiva analisi di proteomica.

Modalità di esame:

Prove in itinere scritte, che possono complementare la valutazione. Esame orale: Allo studente verrà chiesto di presentare un argomento a piacere. Seguiranno due domande specifiche. Lo studente potrà fare uso di diapositive sull'argomento a scelta.

Criteri di valutazione:

Valuteremo le conoscenze acquisite durante il corso inerenti concetti di base sulla teoria dell'evoluzione, la genetica e le leggi di Mendel, biologia molecolare e cellulare di base, caratteristiche di DNA, RNA e proteine, ed espressione genica tessuto-specifica.

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

Le diapositive utilizzate saranno fornite agli studenti.

LAW AND DATA

Titolare: Dott.ssa FIORELLA DAL MONTE

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

MACHINE LEARNING

Titolare: Dott. DAVIDE RIGONI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

MATHEMATICAL CELL BIOLOGY

Titolare: Prof. MORTEN GRAM PEDERSEN

Mutuato da: Scuola Galileiana di Studi Superiori - Classe di Scienze Naturali

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Conoscenza di equazioni differenziali, algebra lineare, teoria di probabilità.

Conoscenze e abilità da acquisire:

Conoscenza di modelli dinamici di biologia cellulare, in particolare di attività neuronale. Conoscenza di analisi di sistemi dinamici deterministici e stocastici. Abilità di fare analisi di sistemi dinamici nonlineari. Abilità di implementazione e simulazione numerica di modelli dinamici in software adeguato. Capacità di valutare la bontà di modelli per loro scopo ("spirito critico")

Attività di apprendimento previste e metodologie di insegnamento:

Circa 2 ore di lezioni e 2 ore di laboratorio (aula informatica) ogni settimana.

Contenuti:

CORSO IN LINGUA INGLESE Attività elettrica in neuroni, modello di Hodgkin-Huxley, bursting. Dinamica stocastica di canali ionici. Sinapsi e gap junctions. Dinamica di calcio. Analisi matematica di sistemi dinamici nonlineari, deterministici e stocastici. Biforcazioni.

Modalità di esame:

Esame scritto (teoria ed esercizi al computer).

Criteria di valutazione:

Saranno valutate le capacità di interpretazione degli risultati matematici e dei simulazioni per la biologia, nonché le capacità di simulare numericamente e analizzare matematicamente i modelli.

Testi di riferimento:

Keener, James; Sneyd, James, Mathematical physiology. New York [etc.]: Springer, 1998 Fall, Christopher P. et al., Computational cell biology. New York [etc.]: Springer, 2000

Eventuali indicazioni sui materiali di studio:

Tutte le dispense delle lezioni e vari file dai laboratori saranno messo disponibili sulla piattaforma Moodle.

NETWORK SCIENCE

Titolare: Prof. TOMASO ERSEGHE

Mutuato da: Scuola Galileiana di Studi Superiori - Classe di Scienze Naturali

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Il corso ha i seguenti prerequisiti: conoscenze in Teoria della Probabilità e Programmazione al Calcolatore in un qualunque linguaggio appropriato per lo studio delle reti (preferibilmente Python, in alternativa MatLab, R, C, o Java); conoscenze in Analisi Matematica e Algebra Lineare; ogni ulteriore conoscenza dei processi di rete in economia, biologia, telecomunicazioni, semantica, etc. può essere utile, così come conoscenze in machine e deep learning o ottimizzazione.

Conoscenze e abilità da acquisire:

Il corso ha le seguenti conoscenze e abilità attese: 1. Apprendere e interpretare criticamente le principali misure usate nell'analisi delle reti 2. Conoscere i principali modelli matematici che descrivono i processi di generazione di una rete 3. Saper riconoscere il livello di importanza dei nodi nella rete 4. Saper identificare comunità (ovvero gruppi coesi), anche parzialmente sovrapposte, usando soluzioni algoritmiche adeguate 5. Valutare il livello di robustezza/coesione di una rete 6. Saper studiare una rete semantica, individuare gli argomenti, e descriverli anche in relazione a metriche socio-psicologiche quali il sentiment 7. Conoscere i principali ambiti di applicazione, anche interdisciplinari, delle tecniche studiate 8. Essere in grado di sintetizzare l'analisi di una rete in un documento professionale 9. Essere in grado di implementare al calcolatore algoritmi atti all'analisi delle reti, anche tramite uso di machine learning

Attività di apprendimento previste e metodologie di insegnamento:

Il modulo prevede: - 18 lezioni frontali che diano una panoramica sulle tematiche e sulle metodologie, che comprendano una carrellata sugli scenari applicativi, e includano un approfondimento matematico sulle tematiche introdotte; - 4 esperienze di laboratorio atte a guidare gli studenti all'uso di programmi al calcolatore per l'analisi di una rete; - 2 lezioni interattive finalizzate a definire i progetti interdisciplinari con gli studenti del corso gemello di Social Network Analytics; - un continuo feedback per la finalizzazione dei progetti.

Contenuti:

Il modulo prevede di coprire i seguenti contenuti: 1. Proprietà basilari di una rete: grafi, matrice di adiacenza, distribuzione del grado dei nodi, connettività, distanza e diametro, coefficiente di clustering. 2. Modelli di rete - Modello Erdos-Renyi; Grafi aleatori; Leggi di potenza e reti scale-free; Fenomeno dello small-world; Hubs; Generazione e espansione di una rete; Modello di Barabasi-Albert; Attaccamento preferenziale. 3. Misure di centralità: Hubs and authorities; PageRank; teletrasporto, ranking su argomenti specifici, misure di prossimità, grado di fiducia; betweenness, closeness, centralità agli autovalori e di Katz. 4. Altre analitiche: omofilia (assortatività), polarizzazione, innovazione, clustering, robustezza, link prediction. 5. Rilevazione unsupervised di comunità: metriche fondamentali di misura (modularità, NMI, Ncut e InfoMap), algoritmi di clustering (Louvain, spectral clustering, InfoMap, stochastic block models, BIGClam, Dendrograms, HDBSCAN) e loro proprietà. 6. Rappresentazione di una rete: funzionalità grafiche di Gephi e R/Python; algoritmi force-directed per la visualizzazione grafica di una rete. 7. Network semantici: costruzione dei network, algoritmi specifici (LDA, NMF, variational autoencoders, the transformer architecture and BERTopic), sentiment analysis. 8. Twitter Lab - Come estrarre una rete semantica da Twitter.

Modalità di esame:

La verifica delle conoscenze e delle abilità attese viene effettuata tramite 1) un ESAME SCRITTO a risposte multiple e 2) lo SVILUPPO DI UN PROGETTO volto a verificare la capacità di applicazione della teoria in contesti anche interdisciplinari, e che richiede: la scelta, la raccolta dei dati e l'analisi di una differente rete per ogni studente; l'implementazione al calcolatore (in un qualsiasi linguaggio di programmazione noto allo studente) degli algoritmi richiesti per l'analisi; la PRESENTAZIONE ORALE dei risultati del progetto. Si prevede un bonus fino a 3 punti per gli studenti frequentanti che partecipino ad un PROGETTO INTERDISCIPLINARE in collaborazione con gli studenti di scienza della comunicazione che frequentano il corso gemello di SOCIAL NETWORK ANALYSIS.

Criteria di valutazione:

I criteri di valutazione con cui verrà effettuata la verifica delle conoscenze e delle abilità attese saranno: 1. Completezza delle conoscenze acquisite 2. Capacità di analisi di una rete attraverso le tecniche proposte 3. Proprietà nella terminologia tecnica usata, sia scritta che orale 4. Originalità e indipendenza nella identificazione della rete oggetto di studio 5. Competenza e coerenza nell'interpretazione del significato delle misure analitiche ottenute 6. Abilità nell'utilizzo degli strumenti informatici nello studio delle misure analitiche di rete.

Testi di riferimento:

Newman, Mark E. J., Networks: an introduction. Oxford: New York, Oxford University Press, 2010 Barabási, Albert-László, Network Science. Cambridge: Cambridge University Press, 2016

Eventuali indicazioni sui materiali di studio:

Tutto il materiale didattico è reso disponibile online @ <http://stem.elearning.unipd.it>

OMICS IN BIOSCIENCE

Titolare: Prof.ssa EMANUELA LEONARDI

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Conoscenze base di informatica, statistica e programmazione.

Conoscenze e abilità da acquisire:

Il corso fornisce le conoscenze biologiche di base sui diversi tipi di dati "omici", ovvero dati ottenuti da indagini sperimentali su larga scala, come esperimenti di genomica, trascrittomica, e proteomica. Inoltre il corso fornirà un'introduzione alle strategie di analisi computazionale dei dati "omici" per lo studio dei meccanismi molecolari sottostanti ai sistemi biologici e alle loro modificazioni/alterazioni. La struttura del corso si articola in moduli dedicati a ciascun tipo di dato, fornendo una conoscenza di base sulle principali tecnologie per la generazione di dati su larga scala, sulla struttura dei vari tipi di dati "omici" e sulle risorse disponibili per la loro acquisizione. Inoltre, saranno esaminati esempi di integrazione dei dati "omici" che hanno condotto a scoperte recenti nel campo della biologia di base e della medicina personalizzata.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso comprende lezioni frontali, esercitazioni pratiche al computer, lo sviluppo di un progetto di gruppo e la sua presentazione con discussione critica. Le esercitazioni forniscono gli strumenti computazionali per condurre un'analisi bioinformatica relativa all'interpretazione dei dati omici riguardanti un problema biologico. La presentazione del progetto consente di valutare l'effettiva comprensione del problema e i punti di forza e debolezze dell'analisi proposta.

Contenuti:

Il corso si compone di tre parti: 1. Genomica ed epigenomica • Geni, regolazione e varianti genetiche • Tecniche di sequenziamento • Processamento dati di sequenziamento • Dati genomici • Banche dati genomiche • Interpretazione dei dati genomici • Applicazioni biologiche 2. Trascrittomica • I trascritti, regolazione temporo-spaziale • Analisi dell'espressione genica • Banche dati di espressione genica • Applicazioni biologiche 3. Proteomica • Proteine, Interazioni e modificazioni • Tecniche di analisi del proteoma • Banche dati • Applicazioni biologiche

Modalità di esame:

L'esame si compone in due parti: 1. Valutazione delle conoscenze pratiche acquisite durante le esercitazioni. 2. Esame finale: domande a risposta aperta e a risposta multipla.

Criteri di valutazione:

Si valuta: • La comprensione di concetti e gli algoritmi presentati a lezione • La capacità di applicare le nozioni fornite a lezione su problemi reali • La capacità critica di saper utilizzare i metodi nei modi più opportuni, scegliendo tra le alternative possibili • La capacità di sviluppare analisi di valore scientifico e riproducibili • La capacità espositiva e di discussione critica

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

Sul sito E-learning vengono resi disponibili molti materiali per il corso. Questi comprendono i lucidi del corso, le registrazioni audio (podcast), e la letteratura citata a lezione.

OPTIMIZATION FOR DATA SCIENCE

Titolare: Prof. FRANCESCO RINALDI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

SEMINARS AND OTHER ACTIVITIES

Titolare: Prof. WOLFGANG ERB

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: II anno, annuale

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: ; 3,00

STATISTICAL LEARNING (C.I.)

Titolare: Prof. ALBERTO ROVERATO

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Moduli del C.I.:

Statistical Learning 1 (Mod. A)
Statistical Learning 2 (Mod. B)

STATISTICAL LEARNING 1 (MOD. A)

Titolare: Prof. ALBERTO ROVERATO

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

STATISTICAL LEARNING 2 (MOD. B)

Titolare: Prof. ALBERTO ROVERATO

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

STOCHASTIC METHODS

Titolare: Prof. MARCO FERRANTE

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

STRUCTURAL BIOINFORMATICS

Titolare: Prof. DAMIANO PIOVESAN

Mutuato da: Laurea magistrale in Computer Science (Ord. 2021)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 32A+16E; 6,00

Prerequisiti:

Conoscenze base di informatica, algoritmi di ottimizzazione e machine learning. Linguaggio di programmazione: Python.

Conoscenze e abilità da acquisire:

L'obiettivo di questo corso è fornire agli studenti una comprensione fondamentale della struttura e della funzione degli organismi viventi a livello molecolare, nonché delle principali metodologie computazionali utilizzate per lo studio degli stessi. Inoltre, il corso mira ad equipaggiare gli studenti con le competenze necessarie per sviluppare in modo autonomo un progetto di ricerca nel campo della bioinformatica strutturale. Gli studenti impareranno a identificare i problemi irrisolti nel campo, valutare lo stato attuale della ricerca, implementare modelli e metodi per affrontare tali problemi e analizzare criticamente i risultati ottenuti. Al termine del corso, gli studenti avranno acquisito le conoscenze e gli strumenti necessari per fare contributi significativi allo studio del materiale vivente, sviluppando al contempo le loro competenze di ricerca e di pensiero critico.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso è strutturato attorno a una combinazione di lezioni, esercitazioni pratiche al computer e lo sviluppo di un progetto di ricerca, che sarà presentato e discusso criticamente. Le esercitazioni pratiche al computer sono progettate per familiarizzare gli studenti con le librerie software utilizzate nei progetti di bioinformatica. Attraverso queste esercitazioni, gli studenti acquisiranno esperienza pratica nell'utilizzo di questi strumenti e svilupperanno le loro competenze nella risoluzione di problemi reali nel campo. Il progetto di ricerca consentirà agli studenti di applicare i concetti e i metodi appresi nel corso a un problema specifico nel campo della bioinformatica. Gli studenti lavoreranno sui loro progetti in gruppo, con il supporto e il feedback del professore, e presenteranno i loro risultati alla classe. La presentazione del progetto richiederà una discussione critica, in cui gli studenti saranno chiamati a identificare i punti di forza e di debolezza dell'implementazione del loro software e a valutare i risultati della loro ricerca. Al termine del corso, gli studenti avranno acquisito una conoscenza approfondita degli strumenti e dei metodi utilizzati nella ricerca di bioinformatica, nonché la capacità di applicare questi concetti a problemi reali e di valutare criticamente i risultati ottenuti.

Contenuti:

Il corso è strutturato attorno ai seguenti argomenti: - Introduzione alla chimica, chimica organica e biochimica - Struttura e funzione delle proteine - Termodinamica del ripiegamento proteico - Metodi sperimentali per la determinazione della struttura - Basi di dati biologiche - Visualizzazione della

struttura con PyMOL - Analisi e manipolazione di strutture proteiche con la libreria BioPython - Relazioni sequenza-struttura nelle proteine e classificazione strutturale - Proteine non globulari, disordine e proteine ripetute - Sovrapposizione e allineamento strutturale - Metodi per la predizione di feature strutturali - Metodi per la predizione della struttura, incluso l'esperimento CASP - Introduzione alla dinamica molecolare

Modalità di esame:

L'esame è suddiviso in tre parti separate, ognuna delle quali deve essere superata per completare con successo il corso. I pesi relativi di ciascuna parte sono indicati tra parentesi di seguito: - Test scritti durante il corso, che consistono in domande teoriche e pratiche - Progetto software - Presentazione del progetto e valutazione critica Per superare l'esame, gli studenti devono completare con successo tutte e tre le parti e ottenere un voto sufficiente in ciascuna di esse. I test scritti durante il corso valuteranno la comprensione dei concetti teorici affrontati nel corso, mentre il progetto software valuterà la capacità di applicare questi concetti a un problema pratico. La presentazione del progetto valuterà invece la capacità di comunicare i risultati della ricerca e di valutare il lavoro dei colleghi.

Criteri di valutazione:

La valutazione finale del corso sarà basata sui seguenti criteri: - Comprensione dei concetti e degli algoritmi presentati in classe - Capacità di applicare i concetti descritti a problemi reali - Abilità di pensiero critico, inclusa la capacità di scegliere i metodi e le alternative più appropriate - Capacità di sviluppare software riutilizzabile implementando nuovi modelli e metodi - Abilità di presentazione e discussione efficaci, inclusa la capacità di valutare criticamente e comunicare i risultati della ricerca.

Testi di riferimento:

Arthur M. Lesk, Introduction to Bioinformatics. : Oxford University Press, 2013 Bruce Alberts, Alexander Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts, and Peter, Molecular Biology of the Cell. : Garland Science, 2014 Donald Voet, Judith G. Voet, and Charlotte W. Pratt, Fundamentals of Biochemistry: Life at the Molecular Level. : Wiley, 2016 Theodore L. Brown, H. Eugene LeMay Jr., and Bruce E. Bursten, Chemistry: The Central Science. : Pearson, 2017 Philip E. Bourne and Helge Weissig, Structural Bioinformatics. : John Wiley & Sons, 2003 Jeremy M. Berg, John L. Tymoczko, and Lubert Stryer, Biochemistry. : W. H. Freeman, 2015

Eventuali indicazioni sui materiali di studio:

Sul sito E-learning vengono resi disponibili molti materiali per il corso. Questi comprendono le slide del corso (appena disponibili) e la letteratura usata per i progetti. Il software utilizzato durante le sessioni pratiche viene condiviso tramite Google Colab Notebooks.

SYSTEMS BIOLOGY

Titolare: Prof. GABRIELE SALES

Mutuato da: Laurea magistrale in Quantitative and Computational Biosciences (Ord. 2024)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Biological Data Analytics

Tipologie didattiche: 40A+16E; 6,00

Prerequisiti:

Aver maturato le conoscenze previste dagli esami del primo anno del Corso di Laurea Magistrale

Conoscenze e abilità da acquisire:

Il corso affronta diversi aspetti relativi ai meccanismi molecolari di regolazione e stabilità dei genomi.

Attività di apprendimento previste e metodologie di insegnamento:

Lezioni frontali e critical reading. Questo approccio consente di approfondire con attività di gruppo gli argomenti di maggiore attualità e favorisce lo spirito critico degli studenti in relazione a risultati innovativi.

Contenuti:

Introduzione alla Biologia dei Sistemi. (0.25 CFU) Basi su derivate, integrali ed equazioni differenziali (0.25 CFU) Modellazione matematica. (0.5 CFU) Modelli statici di reti. (0.5 CFU) Modelli di Markov. (0.5 CFU) Mutua Informazione, Relevance Networks e reti bayesiane. (0.5 CFU) La matematica dei sistemi biologici. (0.5 CFU) Stima dei parametri a partire da dati rumorosi: ricerche a griglia, hill climbing ed algoritmi genetici. (0.5 CFU) Sistemi di segnale. (0.5 CFU) Sistemi di popolazioni. (0.5 CFU) Simulazione del modello SIR. (0.5 CFU)

Modalità di esame:

La verifica delle conoscenze acquisite avviene attraverso una prova scritta con 4 domande aperte, volte ad evidenziare le conoscenze, il lessico scientifico, la capacità di sintesi e di discussione critica acquisite durante il corso. Questa prova è basata sui temi trattati e discussi a lezione.

Criteri di valutazione:

Comprensione degli argomenti e capacità critica di collegamento delle conoscenze acquisite, tra loro e con i concetti fondamentali della biologia molecolare

Testi di riferimento:

Eberhard O. Voit, A First Course in Systems Biology. : Garland Science, 2012

Eventuali indicazioni sui materiali di studio:

Articoli e capitoli di libro, su indicazione del docente

Curriculum: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

ADVANCED TOPICS IN DATA SCIENCE

Titolare: Prof. LAMBERTO BALLAN

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, annuale

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 24A; 3,00

Prerequisiti:

Nessuno.

Conoscenze e abilità da acquisire:

Comprendere metodi e modelli avanzati per la Data Science.

Attività di apprendimento previste e metodologie di insegnamento:

- Il corso si baserà su lezioni frontali; - Il docente utilizzerà la lavagna e le slides; - risorse (e.g., slide) verranno rese disponibili sulla piattaforma moodle.

Contenuti:

Il corso analizzerà modelli e metodi avanzati per la Data Science.

Modalità di esame:

progetto e/o homework.

Criteri di valutazione:

La valutazione della preparazione dello studente si baserà: - sulla comprensione degli argomenti svolti in aula; - sulla capacità di utilizzare in maniera autonoma e consapevole i modelli e le metodologie risolutive proposte.

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

Dispense o slide fornite dal docente.

BUSINESS ECONOMIC AND FINANCIAL DATA

Titolare: Prof.ssa MARIANGELA GUIDOLIN

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Program of Statistical Learning course

Conoscenze e abilità da acquisire:

The course aims at introducing the students to the main statistical features and concepts underlying the analysis of data collected over time, as well as providing the basic statistical solutions to analyse such data in economic, financial and business settings.

Attività di apprendimento previste e metodologie di insegnamento:

The course is composed of frontal lectures and technical-practical lessons with R

Contenuti:

-General introduction to business, economic and financial data --Preliminary concepts and illustrative examples --Moving beyond the linear model --Linear regression model: main ideas and assumptions --Nonlinear regression models for new product growth - Beyond linearity: regression splines, local regression, generalized additive models (GAM) - Tree-based methods: Regression trees, Bagging, Boosting -Time series analysis: Exponential Smoothing and ARIMA models

Modalità di esame:

--Practical exam --Oral exam

Criteri di valutazione:

Students will be evaluated on their knowledge of models and techniques proposed to analyse economic, financial or business data and their ability to apply them to real cases with a critical attitude.

Testi di riferimento:

Azzalini, A., Scarpa, B., Data analysis and data mining: An introduction.. USA: OUP, 2012 Hyndman, R. J., Athanasopoulos, G., Forecasting: principles and practice. : OTexts., 2018 James, Gareth; Witten, Daniela; Hastie, Trevor; Tibshirani, Robert, An introduction to statistical learning: with applications in R. New York: Springer, 2013

Eventuali indicazioni sui materiali di studio:

Notes prepared by instructor will be made available throughout the course.

COGNITION AND COMPUTATION

Titolare: Prof. MARCO ZORZI

Mutuato da: Laurea magistrale in Cybersecurity (Ord. 2020)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Il corso richiede conoscenze di base sull'apprendimento automatico e sulla teoria della probabilità. Sono inoltre richieste capacità di programmazione in Python per svolgere il mini-progetto individuale richiesto per l'esame. La familiarità con concetti base delle scienze cognitive può facilitare la comprensione dei temi trattati.

Conoscenze e abilità da acquisire:

Il corso fornisce conoscenze sui principali approcci computazionali utilizzati per modellizzare le funzioni cognitive, dalle reti neurali artificiali ai modelli probabilistici. Queste conoscenze sono rilevanti sia per la comprensione del funzionamento della mente che per lo sviluppo e la valutazione dei moderni sistemi di intelligenza artificiale. La discussione teorica dei diversi approcci verrà affiancata da esempi concreti di applicazione (es. visione, linguaggio, ragionamento).

Attività di apprendimento previste e metodologie di insegnamento:

L'insegnamento è basato su lezioni frontali in cui vengono trattati gli argomenti teorici. Verranno utilizzate tecniche di apprendimento interattivo, come le discussioni interattive su domande aperte, per promuovere la riflessione critica sui concetti discussi. Il corso include alcune esercitazioni (in aula informatica) con simulazioni al computer.

Contenuti:

1. Introduzione: modellizzazione computazionale e matematica nelle scienze cognitive e nelle neuroscienze cognitive. Rassegna degli approcci simbolici, emergentisti e probabilistici alla simulazione della cognizione umana. 2. Modelli probabilistici della cognizione: concetti base su inferenza Bayesiana e modelli grafici probabilistici; apprendimento induttivo; programmazione probabilistica. 3. Modelli connessionisti della cognizione: concetti base sulla computazione neurale; apprendimento nelle reti neurali; architetture per deep learning. 4. Casi di studio: visione e acquisizione di concetti; elaborazione del linguaggio; ragionamento.

Modalità di esame:

L'esame consisterà in una prova scritta con circa 20 domande a scelta multipla (durata massima 30 minuti). Ogni studente dovrà inoltre consegnare entro il giorno precedente l'esame scritto un Python notebook del progetto assegnato durante il corso.

Criteri di valutazione:

La valutazione sarà basata sulla comprensione degli argomenti trattati nel corso e sull'acquisizione dei concetti e delle metodologie proposte.

Testi di riferimento:

Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron, Deep learning. London, England; Cambridge, Mass: The MIT Press, 2016 Koller, Daphne; Friedman, Nir, Probabilistic Graphical Models. Cambridge: MIT Press, 2009 Russell, Stuart; Norvig, Peter, Artificial Intelligence. Harlow: Pearson Education UK, 2013

Eventuali indicazioni sui materiali di studio:

Tutti gli argomenti verranno trattati durante le lezioni. Le slides delle lezioni saranno disponibili sulla piattaforma di e-learning Moodle. Gli appunti degli studenti dovranno essere integrati dai libri di testo e da altro materiale (articoli scientifici) forniti dal docente sulla piattaforma di e-learning.

COGNITIVE, BEHAVIORAL AND SOCIAL DATA

Titolare: Dott.ssa MERYLIN MONARO

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Nozioni di apprendimento automatico.

Conoscenze e abilità da acquisire:

Alla fine del corso gli studenti saranno in grado di comprendere problemi complessi nelle scienze cognitive, sociali e comportamentali, di scegliere le metodologie più appropriate per estrarre informazioni dai dati, e di integrare le conoscenze di data science con aspetti riguardanti le scienze sociali, il cervello, la mente e il comportamento. Verranno inoltre acquisiti: - I concetti di base di psicologia cognitiva, psicologia sociale e scienze del comportamento. - Gli strumenti e le metodologie dell'analisi dei dati cognitivi, comportamentali e sociali. - Abilità pratiche di analisi dei dati applicata a problemi cognitivi, comportamentali e sociali.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso prevede sia lezioni frontali che attività in piccoli gruppi.

Contenuti:

Lo scopo del corso è di fornire una panoramica di applicazioni concrete della Data Science alle scienze comportamentali, cognitive, sociali e alle neuroscienze. Il corso fornisce le basi dei metodi per analizzare dati comportamentali, cognitivi, e relativi a funzionalità e struttura del cervello. La panoramica fornita includerà esempi di recenti applicazioni, selezionati anche a seconda degli interessi degli studenti. Verranno discussi i limiti dello stato dell'arte e le direzioni di sviluppo future. I contenuti saranno i seguenti. • Concetti di base sul cervello umano e il suo funzionamento cognitivo (anatomia cerebrale, attenzione, memoria, apprendimento, linguaggio ecc.) e relative misure • Concetti di base di psicologia sociale e comportamento sociale (preferenze, giudizio, identità di gruppo, ecc.) e relative misure • Misure comportamentali e come ottenerle (es. RT); misure di comportamento implicite ed esplicite (es. la IAT) • Estrarre e predire informazioni dal comportamento (es. lie detection, predizione di "malicious behavior" dall'attività sui social networks, fake online reviews, security, ecc.) • Misure psicofisiologiche e come ottenerle (es. HR variability, SCR, espressioni facciali, EEG, fMRI, etc.) • Estrarre e predire informazioni dalle misure psicofisiologiche • Estrarre e predire informazioni dalle attività cerebrali: "mind reading applications" (es. psychopathology detection, ricostruzione di esperienze visive dall'attività cerebrale, brain computer interface devices, ecc.) • Applicazioni al marketing di dati sociali e comportamentali (es. skill assessment and prediction, psychology of taxes, predicting preferences and personality from social networks activity, sentiment analysis, ecc.) • Questioni legate all'applicazione di apprendimento automatico nella ricerca comportamentale (es. il problema della riproducibilità)

Modalità di esame:

Esame scritto con domande chiuse e aperte.

Criteri di valutazione:

Verrà valutata la conoscenza degli argomenti proposti durante le lezioni, l'acquisizione dei concetti e metodologie, e l'abilità di applicarli.

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

Materiale presentato dai docenti durante il corso, che sarà messo a disposizione degli studenti su Moodle. Articoli scientifici inerenti al contenuto delle lezioni, che saranno messi a disposizione degli studenti su Moodle.

DEEP LEARNING (MATICOLE DISPARI)

Titolare: Prof. ALESSANDRO SPERDUTI

Mutuato da: Laurea magistrale in Computer Science (Ord. 2021)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

E' opportuno possedere le conoscenze di base relative al Calcolo delle Probabilità, alla Programmazione e agli Algoritmi.

Conoscenze e abilità da acquisire:

L'insegnamento introduce i concetti di base relativi al Deep Learning, cioè all'apprendimento automatico tramite reti neurali. Verranno richiamati i concetti matematici necessari per una piena comprensione della materia. Si tratteranno le reti neurali feedforward deep e le relative tecniche di regolarizzazione e di ottimizzazione dell'apprendimento. Verranno introdotti i concetti di base relativi alle reti convolutive. Per quanto riguarda il trattamento di sequenze, saranno presentate le reti neurali ricorrenti, e il Transformer. Infine si tratteranno autoencoder e modelli generativi deep. Inoltre, gli studenti dovranno consegnare homework pratici che coprono i contenuti dell'insegnamento.

Attività di apprendimento previste e metodologie di insegnamento:

L'insegnamento prevede lezioni frontali.

Contenuti:

La tematiche dell'insegnamento saranno le seguenti: - Introduzione ai contenuti dell'insegnamento; - Reti Neurali Feedforward profonde (deep); - Regolarizzazione per l'apprendimento deep; - Ottimizzazione per l'apprendimento di modelli deep; - Concetti di base per reti neurali convolutive; - Reti neurali ricorrenti e Transformers per la modellazione di sequenze; - Autoencoder; - Modelli generativi deep;

Modalità di esame:

Lo studente deve superare un esame scritto. Per essere ammesso all'esame lo studente deve aver consegnato ed ottenuto valutazione positiva tutti gli homework previsti dall'insegnamento.

Criteri di valutazione:

La valutazione dello studente si basa su una verifica dell'apprendimento dei concetti di base introdotti durante il corso e sulla capacità di analisi dello studente.

Testi di riferimento:

Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron, Deep Learning. Cambridge: MA, MIT Press, 2016

Eventuali indicazioni sui materiali di studio:

Materiale aggiuntivo sarà disponibile sul sito e-learning del corso.

DEEP LEARNING (MATICOLE PARI)

Titolare: Prof. ALESSANDRO SPERDUTI

Mutuato da: Laurea magistrale in Computer Science (Ord. 2021)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

E' opportuno possedere le conoscenze di base relative al Calcolo delle Probabilità, alla Programmazione e agli Algoritmi.

Conoscenze e abilità da acquisire:

L'insegnamento introduce i concetti di base relativi al Deep Learning, cioè all'apprendimento automatico tramite reti neurali. Verranno richiamati i concetti matematici necessari per una piena comprensione della materia. Si tratteranno le reti neurali feedforward deep e le relative tecniche di regolarizzazione e di ottimizzazione dell'apprendimento. Verranno introdotti i concetti di base relativi alle reti convolutive. Per quanto riguarda il trattamento di sequenze, saranno presentate le reti neurali ricorrenti, e il Transformer. Infine si tratteranno autoencoder e modelli generativi deep. Inoltre, gli studenti dovranno consegnare homework pratici che coprono i contenuti dell'insegnamento.

Attività di apprendimento previste e metodologie di insegnamento:

L'insegnamento prevede lezioni frontali.

Contenuti:

La tematiche dell'insegnamento saranno le seguenti: - Introduzione ai contenuti dell'insegnamento; - Reti Neurali Feedforward profonde (deep); - Regolarizzazione per l'apprendimento deep; - Ottimizzazione per l'apprendimento di modelli deep; - Concetti di base per reti neurali convolutive; - Reti neurali ricorrenti e Transformers per la modellazione di sequenze; - Autoencoder; - Modelli generativi deep.

Modalità di esame:

Lo studente deve superare un esame scritto. Per essere ammesso all'esame lo studente deve aver consegnato ed ottenuto valutazione positiva tutti gli homework previsti dall'insegnamento.

Criteri di valutazione:

La valutazione dello studente si basa su una verifica dell'apprendimento dei concetti di base introdotti durante il corso e sulla capacità di analisi dello studente.

Testi di riferimento:

Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron, Deep Learning. Cambridge: MA, MIT Press, 2016

Eventuali indicazioni sui materiali di studio:

Materiale aggiuntivo sarà disponibile sul sito e-learning del corso.

ENGLISH LANGUAGE B2 (PRODUCTIVE SKILLS)

Titolare: Prof. WOLFGANG ERB

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, annuale

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: ; 3,00

FINAL EXAMINATION

Titolare: da definire

Periodo: Il anno, annuale

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: ; 15,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

FINANCIAL MATHEMATICS FOR DATA SCIENCE

Titolare: Prof. CLAUDIO FONTANA

Mutuato da: Laurea magistrale in Computational Finance (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Solide conoscenze di probabilità, calcolo integrale e differenziale. Una conoscenza dei concetti di base di economia finanziaria e dei mercati finanziari è consigliata.

Conoscenze e abilità da acquisire:

ABILITA COGNITIVE Al termine del corso gli studenti saranno in grado di: C1. Discutere il funzionamento dei principali tipi di derivati finanziari; C2. Spiegare e analizzare modelli stocastici per i mercati finanziari; C3. Comprendere le tecniche di valutazione dei derivati finanziari; C4. Analizzare i rischi dei derivati finanziari. **ABILITÀ PRATICHE** Gli studenti saranno in grado di: P1. Calcolare il prezzo di derivati finanziari; P2. Gestire il rischio di derivati finanziari; P3. Costruire strategie di trading e portafogli di hedging. P4. Sviluppare modelli di valutazione. P5. Implementare metodi numerici per la soluzione di problemi di valutazione e hedging. **COMPETENZE TRASVERSALI** Gli studenti svilupperanno: T1. Abilità matematiche; T2. Abilità computazionali; T3. Capacità di problem solving; T4. Capacità analitiche.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso prevede: • Lezioni frontali; • Esercitazioni; • Discussione di esempi e case studies.

Contenuti:

Il corso tratterà gli argomenti seguenti: - Introduzione ai contratti derivati - Il modello binomiale: valutazione, copertura, probabilità risk-neutral - Nozioni di base sui processi stocastici a tempo continuo - Calcolo stocastico e equazioni differenziali stocastiche - Il modello Black-Scholes: valutazione, copertura, probabilità risk-neutral, lettere greche - Metodi Monte-Carlo e tecniche di riduzione della varianza - Modelli a volatilità stocastica e locale - Modelli stocastici per i tassi di interesse - Cenni ai modelli stocastici per il rischio di credito

Modalità di esame:

Per gli studenti di Data Science: esame scritto e valutazione di homework. Per gli studenti di Computational Finance: esame scritto, valutazione di homework e progetto.

Criteri di valutazione:

Per gli studenti di Data Science: esame scritto, fino a 5 punti per homework risolti correttamente. Per gli studenti di Computational Finance: 2/3 esame scritto, 1/3 progetto, fino a 5 punti per homework risolti correttamente. Nell'esame scritto e negli homework, saranno valutate le conoscenze teoriche possedute dallo studente, la capacità di risolvere esercizi (calcolare prezzi e strategie di copertura di derivati finanziari, risolvere esercizi di calcolo stocastico) e di analizzare strategie di investimento che coinvolgono derivati finanziari. Saranno valutate le abilità C1-C4, P1-P3 e T1-T4. Nel progetto, sarà valutata la capacità di analizzare il problema e fornire una soluzione numerica adeguata. Tutte le abilità sopra descritte verranno valutate.

Testi di riferimento:

Hull, John C., Options, futures, and other derivatives. Harlow: Pearson, 2022 Seydel, Rüdiger U, Tools for Computational Finance. London: Springer London, 2017 Bjork, Tomas, Arbitrage Theory in Continuous Time. Oxford: Oxford University Press, 2009

Eventuali indicazioni sui materiali di studio:

Slides delle lezioni (rese disponibili sulla pagina Moodle del corso), appunti del docente, testi di riferimento indicati in bibliografia.

FUNDAMENTALS OF INFORMATION SYSTEMS
--

Titolare: Prof. GIORGIO MARIA DI NUNZIO

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 96A; 12,00

Prerequisiti:

The student should have basic knowledge of computer programming and problem solving skills.

Conoscenze e abilità da acquisire:

The aim of this class is to teach the concepts, methods, and technologies which any modern data scientist should master. In particular, the focus of this class is on the processing/storing of data and big data, which also involves elements of algorithmic methods and data structures. The ability of processing data effectively and efficiently will be gained using Python, which is one of the reference programming language for data scientists. Ultimately, students will acquire coding skills to collect, clean, visualize, and analyse data efficiently, and, more generally to tackle any data science/machine learning task. Concerning storage, the basics of relational databases are introduced, followed by a review of non-relational solutions typically adopted for big data. The part of the course covering algorithmic methods will first concentrate on the main data structures and their efficient implementation. Attention will then shift to the fundamental algorithmic paradigms for problem solving and their applicability through the discussion of relevant case studies.

Attività di apprendimento previste e metodologie di insegnamento:

The course consists of lectures. Students will have access to online material for self-evaluation of programming skills

Contenuti:

The course is structured into 3 submodules: - Python Programming (for Data Science) This submodule provides students with the foundational coding skills they need as data scientists. First, the basics of the Python programming language are covered (i.e., built-in data types, functions, I/O, etc.) along with the environment which is used throughout the class (i.e., Jupyter Notebook). Afterwards, students will dig into a set of the most up-to-date data science Python packages; those are: numpy/scipy (for numerical/scientific computing), pandas (for data manipulation), matplotlib/seaborn (for data visualization), and finally scikit-learn (for learning from data). - Databases This submodule is dedicated to data storage, and it covers the following topics: Relational databases, Logical and Physical Design of a Relational Database. SQL Language: Data Definition and Data Manipulation Language, Database Query The PostgreSQL database: Creation and Definition of a Database, SQL Queries. Non Relational databases, graph databases, Cypher query language. Neo4J database: Creation and Definition of a Database, Graph Query Language. - Algorithmic Methods: Preliminaries: definition of problem, instance, solution, algorithm. Models of computation. Analysis of algorithms: correctness and running time. Asymptotic analysis. Basic data structures: lists, stacks, queues. Trees and their properties. Dictionaries and their implementation. Priority queues. Graphs: representation of graphs. Basic properties. Graph searches and applications. Divide and Conquer paradigm: the use of recursion. Case study: sorting. Eventually, at the end all the modules, students will be able to implement all the stages of a typical machine learning pipeline: from collecting data to building predictive models for solving efficiently a data analysis/prediction problem.

Modalità di esame:

Written exam.

Criteri di valutazione:

The written exam will be evaluated on the basis of the following criteria: i) student's knowledge of the concepts, methods, and technologies at the basis of the topics covered in the course; ii) student's capacity for synthesis, clarity, and abstraction.

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

Slides presented during the lectures are made available to students as reference material. The Python submodule will follow the book "Whirlwind tour of Python" and "Python Data Science Handbook. Both are freely available at <https://jakevdp.github.io/WhirlwindTourOfPython/> <https://jakevdp.github.io/PythonDataScienceHandbook/> The database submodule will follow the book "Database Systems - Concepts, Languages and Architectures" by Paolo Atzeni et al., which is freely available at <http://dbbook.dia.uniroma3.it/> and other online material For the algorithmic part we will consider parts of the book: Michael T. Goodrich, Roberto Tamassia, Michael H. Goldwasser, Data Structures and Algorithms in Python, John Wiley & Sons Inc, 2013

HIGH DIMENSIONAL PROBABILITY FOR DATA SCIENCE

Titolare: Prof. ALBERTO CHIARINI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Un corso di Probabilità. Conoscenze elementari di Processi Stocastici, Algebra Lineare e Geometria. Sarà utile avere qualche familiarità con le nozioni di spazio metrico, normato e di Hilbert.

Conoscenze e abilità da acquisire:

Conoscenza approfondita degli argomenti visti durante il corso. Capacità di risolvere problemi anche complessi e di leggere articoli di ricerca sugli argomenti trattati nel corso.

Attività di apprendimento previste e metodologie di insegnamento:

Lezioni frontali. Verranno assegnati homework e progetti.

Contenuti:

Il corso tenterà di coprire i seguenti argomenti. - Preliminaries on random variables. Classical inequalities and limit theorems. - Concentration of sums of independent random variables: Hoeffding, Chernoff and Bernstein inequalities, sub-Gaussian and sub-exponential distributions. Applications to random graphs. - Random vectors in high dimension: concentration of the norm, covariance matrices and principal component analysis, high dimensional distributions, sub-Gaussian distributions in higher dimensions. Applications: Grothendieck's Inequality and semidefinite programming and maximum cut for graphs. - Non-asymptotic analysis of random matrices: nets, covering and packing numbers, bounds on sub-Gaussian matrices, covariance estimation and clustering. Applications to error correcting codes, community detection in networks and covariance estimation and clustering. - Concentration of Lipschitz functions on the sphere, Johnson-Lindenstrauss theorem, matrix Bernstein inequality, community detection in sparse networks. - Random processes: basic concepts, Slepian's inequality, bounds on Gaussian matrices, Sudakov's minoration inequality, Gaussian width, random projections of sets. - Chaining: Dudley's inequality, empirical processes, Vapnik-Chervonenkis dimension with applications to statistical learning theory.

Modalità di esame:

La valutazione sarà basata sulla valutazione di esercizi per casa assegnati settimanalmente e su un orale. In alternativa, è possibile lo sviluppo e la presentazione di un progetto.

Criteri di valutazione:

Alla fine del corso gli studenti conosceranno le tecniche probabilistiche per lo studio di problemi con molti gradi di libertà e le loro applicazioni ai big data. Gli studenti di devono essere capaci di risolvere esercizi di difficoltà adeguata.

Testi di riferimento:

Roman Vershynin, High Dimensional Probability for Mathematicians and Data Scientists. : Cambridge University Press, 2018 Ramon Van Handel, Probability in High Dimension. : Princeton University, 2016

Eventuali indicazioni sui materiali di studio:

Tutti gli argomenti d'esame vengono illustrati a lezione. Materiale addizionale (esercizi, appunti del docente) saranno disponibile su Moodle.

HUMAN COMPUTER INTERACTION

Titolare: Prof. LUCIANO GAMBERINI

Mutuato da: Scuola Galileiana di Studi Superiori - Classe di Scienze Naturali

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 42A; 6,00

Prerequisiti:

Non sono richiesti particolari prerequisiti. Per gli studenti che parlano italiano, si suggerisce di frequentare contemporaneamente il laboratorio di INTERACTION DESIGN progettato per mettere ulteriormente in pratica quanto appreso in questo corso.

Conoscenze e abilità da acquisire:

Il corso offre la possibilità di acquisire conoscenze teoriche, metodi di ricerca e tecniche innovative per lo studio, la progettazione e la valutazione dell'interazione tra le persone e le tecnologie. Tali conoscenze sono utili per rendere l'interazione persona-macchina efficace ed efficiente e l'esperienza d'uso semplice, piacevole e complessivamente soddisfacente per l'utente. Le competenze che si acquisiranno interesseranno quindi i domini di conoscenza dell'interazione persona-computer (HCI) e dell'ergonomia cognitiva; in dettaglio si acquisiranno competenze negli ambiti: - del design centrato sull'utente - dei principi di base dell'ergonomia cognitiva - della valutazione dell'esperienza dell'utente e dell'usabilità dei prodotti - della comunicazione visiva e della visualizzazione dei dati - dell'accessibilità e del design universale (es: design for older adults) - del social computing e dell'ergonomia sociale

Attività di apprendimento previste e metodologie di insegnamento:

Lezioni tradizionali e interattive (presentazioni studenti, multimedia, risorse on-line) sugli aspetti teorici della disciplina saranno intervallati da laboratori didattici in cui si sperimenteranno i metodi e le tecniche appresi durante il corso. Lavori individuali e di gruppo tramite il design e lo sviluppo di prototipi di interfacce e sistemi interattivi permetteranno allo studente di acquisire competenze specifiche e pratiche. Sono benvenute, ma non sono obbligatorie particolari precedenti competenze tecniche o informatiche.

Contenuti:

Seguendo il libro si analizzeranno i seguenti argomenti: 1 What is Interaction Design? 2 The Process of Interaction Design 3 Conceptualizing Interaction 4 Cognitive Aspects 5 Social Interaction 6 Emotional Interaction 7 Interfaces 8 Data Gathering 9 Data Analysis, Interpretation, and Presentation 10 Data at Scale 11 Discovering Requirements 12 Design, Prototyping, and Construction 13 Interaction Design in Practice 14 Introducing Evaluation 15 Evaluation Studies: From Controlled to Natural Settings 16 Evaluation: Inspections, Analytics, and Models Durante le lezioni verranno discussi e sperimentati praticamente metodi di ricerca e tecniche per la progettazione e la valutazione di sistemi interattivi.

Modalità di esame:

L'esame sarà scritto, della durata di 90 minuti, con eventuale integrazione orale. Ci saranno indicativamente 6-12 domande chiuse (che prevedono risposte puntuali e specifiche) e 1-3 aperte (che richiedono riflessione ed elaborazione), a seconda della complessità delle medesime. Modalità alternative potranno essere previste per i frequentanti.

Criteri di valutazione:

In generale, ai fini di una valutazione positiva, è necessario che lo studente sappia argomentare in modo corretto e pertinente, durante l'esame, le tematiche trattate a lezione.

Testi di riferimento:

Rogers, Yvonne; Sharp, Helen; Preece, Jennifer; Rogers, Yvonne, Interaction design beyond human-computer interaction. Hoboken, NJ: Wiley, 2023

Eventuali indicazioni sui materiali di studio:

FREQUENTANTI INTERACTION DESIGN (6th ed.)- di Helen Sharp, Jenny Preece e Yvonne Rogers è il libro di testo. Materiali didattici saranno a disposizione a lezione e su moodle per tutti gli studenti frequentanti e sostituiranno parti del libro. NON FREQUENTANTI INTERACTION DESIGN (6th ed.)- di Helen Sharp, Jenny Preece e Yvonne Rogers è l'unico materiale da utilizzare per questo corso se non lo si frequenta.

INTERNSHIP

Titolare: Prof. TOMASO ERSEGHE

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, annuale

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: ; 18,00

KNOWLEDGE REPRESENTATION AND LEARNING

Titolare: Prof. ROBERTO CONFALONIERI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

E' opportuno che gli studenti abbiano una conoscenza di base di logica matematica e di statistica.

Conoscenze e abilità da acquisire:

Abilità cognitive. Al termine del corso, gli studenti saranno in grado di: - Comprendere i principi delle logiche per la rappresentazione e il ragionamento della conoscenza. - Comprendere i principi dell'apprendimento statistico relazionale. - Comprendere i principi dell'integrazione neuro-simbolica, ovvero come combinare logiche e apprendimento statistico relazionale per costruire sistemi di apprendimento e ragionamento in domini ibridi. Abilità pratiche. Gli studenti saranno in grado di: - Modellare problemi in logiche come la Logica Proposizionale, la Logica del Primo Ordine e la Logica Fuzzy. - Utilizzare algoritmi di rappresentazione della conoscenza e di ragionamento per trovare i modelli di una teoria logica. - Modellare problemi utilizzando approcci di apprendimento statistico relazionale come modelli grafici, reti logiche di Markov, Prolog probabilistico e reti logiche tensoriali. Abilità trasversali. Gli studenti svilupperanno capacità di pensiero analitico, capacità di problem solving e capacità di pensiero sistemico-logico.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso prevede 48 ore di lezione. Le lezioni saranno supportate da esercitazioni pratiche.

Contenuti:

(A) Logiche per la rappresentazione della conoscenza: (A.i) introduzione alle logiche proposizionali, sintassi, semantica, procedura decisionale. Soddisfabilità, soddisfabilità ponderata e migliore soddisfabilità. (A.ii) Logica del primo ordine, sintassi, semantica, risoluzione e unificazione. (A.iii) Logiche fuzzy, sintassi, semantica e ragionamento. (B) Apprendimento statistico relazionale: (B.i) Modelli grafici (B.ii) Reti logiche di Markov (B.iii) Prolog probabilistico, (B.iv) Reti logiche tensoriali

Modalità di esame:

Le conoscenze e le competenze degli studenti saranno valutate attraverso: - Un esame scritto: agli studenti verrà chiesto di risolvere esercizi simili a quelli illustrati nelle lezioni. - Un project work: agli studenti verrà chiesto di implementare e valutare uno degli algoritmi presentati nelle lezioni o di utilizzare uno strumento di inferenza logica/probabilistica per risolvere un problema pratico.

Criteri di valutazione:

Gli studenti saranno valutati in base alla conoscenza critica degli argomenti del corso. I criteri di valutazione consistono in: 50% Esame scritto: capacità di presentare e applicare il materiale studiato. 50% Lavoro di progetto: capacità di applicare il contenuto del corso a problemi pratici.

Testi di riferimento:

Russell, Stuart J.; Norvig, Peter; Russell, Stuart J., Artificial intelligence a modern approach. Harlow: Pearson Education, 2022 Stuart Russell, Unifying logic and probability. : Communications of the ACM, 2015

Eventuali indicazioni sui materiali di studio:

Il materiale di studio consiste in: - diapositive delle lezioni, - appunti delle lezioni.

LAW AND DATA

Titolare: Dott.ssa FIORELLA DAL MONTE

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 48A; 6,00

MACHINE LEARNING

Titolare: Dott. DAVIDE RIGONI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Sono richieste conoscenze di base di Analisi Matematica, Probabilità e Statistica, Algebra Lineare, e di Programmazione.

Conoscenze e abilità da acquisire:

Il corso presenta i concetti e gli algoritmi fondamentali dell'Apprendimento Automatico, cioè quella classe di tecniche ed algoritmi che a partire dai dati consentono di acquisire nuova conoscenza. Questi sono particolarmente utili per tutti quei problemi per cui è impossibile, o molto difficile, pervenire ad una formalizzazione utilizzabile per la definizione di una soluzione algoritmica ad hoc. Esempi di tali problemi sono i compiti percettivi, come il riconoscimento visivo o il riconoscimento del parlato, e problemi in cui i dati sono affetti da rumore o sono incompleti. Sono previste varie esercitazioni in laboratorio informatico, al fine di consentire allo studente di sperimentare le conoscenze acquisite mediante la loro applicazione a piccoli esempi pratici. Il corso insegna inoltre le competenze necessarie per applicare e valutare modelli di apprendimento automatico in contesti applicativi reali.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso consiste in lezioni e esercizi in laboratorio informatico. Gli esercizi in laboratorio informatico consentono agli studenti di sperimentare, in diversi scenari operativi, le tecniche introdotte a lezione. Gli studenti possono così verificare sperimentalmente i concetti appresi in classe, ed acquisire la capacità di applicarli criticamente di fronte ad un problema concreto.

Contenuti:

Il corso tratta gli argomenti elencati di seguito: - Introduzione Intelligenza Artificiale ed Apprendimento Automatico, quando e perché utilizzare tecniche di Apprendimento Automatico; I diversi paradigmi di Apprendimento Automatico, caratteristiche ed applicazioni; Gli ingredienti Fondamentali dell'Apprendimento Automatico. - Apprendimento Supervisionato: le fondamenta ed i modelli base Modelli di Regressione Lineare; spazio delle ipotesi, rappresentazione e definizione della funzione di costo; ottimizzazione e discesa del gradiente. Modelli di Classificazione Lineare; Regressione Logistica; Regolarizzazione e model selection. - Bontà di un modello e sua valutazione Bias-Variance Tradeoff e come affrontare problemi di overfitting ed underfitting; minimizzazione del rischio e teoria dell'apprendimento; misure di valutazione delle performance, esempi ed applicazioni; come effettuare debugging e diagnosticare il comportamento di un sistema di Apprendimento Automatico. - Apprendimento Supervisionato: reti neurali e modelli avanzati Reti Neurali Artificiali; perceptron, reti multistrato e cenni di deep learning; apprendimento dei parametri in reti neurali, backpropagation e discesa di gradiente. Support Vector Machines; metodi kernel e classificazione non lineare. Paradigmi "alternativi": metodi non-parametrici (k-NN) e loro applicazioni; alberi di decisione e random forest. - Apprendimento non Supervisionato Clustering: KMeans e cenni agli approcci principali; PCA e riduzione di dimensionalità. - Introduzione ai sistemi di Raccomandazione e alla loro applicazione

Modalità di esame:

L'esame verrà svolto con una prova scritta.

Criteri di valutazione:

La valutazione dello studente si basa sull'accertamento dell'apprendimento degli argomenti trattati nel corso e sulle sue capacità di analisi.

Testi di riferimento:

Bishop, Christopher M., Pattern recognition and machine learning. New York: Springer, 2006

Eventuali indicazioni sui materiali di studio:

Le presentazioni mostrate durante le lezioni e link di approfondimento sono resi disponibili su Moodle come materiale di riferimento.

MACHINE LEARNING FOR HUMAN DATA

Titolare: Prof. MICHELE ROSSI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Il corso richiede conoscenze in Teoria della Probabilità (variabili aleatorie, probabilità condizionata e formula di Bayes, distribuzioni di probabilità discrete e continue) e un minimo di conoscenza di linguaggi di programmazione (MatLab, Python). Anche se non strettamente necessarie, conoscenze nell'ambito di algebra lineare (es, spazi vettoriali e decomposizione ai valori singolari) e dell'analisi numerica dei segnali (trasformate di Fourier discrete) sono utili. Il Professore effettuerà dei ripassi per le tecniche e teorie di cui sopra, qualora necessario. Materiale e referenze per il ripasso personale da parte dello studente di tali tecniche verranno altresì forniti. Inoltre, anche se conoscenze pregresse sono sicuramente utili per lo studente, il corso è concepito in modo da non dipendere criticamente da esse. Benchè non obbligatorio, un esame di base di "Machine Learning" è utile per una migliore comprensione del corso.

Conoscenze e abilità da acquisire:

1. Conoscere i principali algoritmi per la classificazione di dati multidimensionali, i loro pro e contro, le metriche per la loro valutazione 2. Conoscere le principali tecniche per l'apprendimento non-supervisionato (vector quantization), le loro prestazioni, i vantaggi e il loro utilizzo in seno a problemi reali nell'ambito dei biosignali 3. Conoscere le principali tecniche di modellizzazione di serie temporali multivariate e il loro utilizzo per la soluzione di problemi reali 4. Conoscere i principi e le tecniche dell'apprendimento supervisionato con particolare riferimento alle reti neurali (feed forward, convoluzionali e ricorrenti), la loro programmazione in Python, e il loro utilizzo in seno a problemi reali 5. Conoscere i principali ambiti di applicazione, delle tecniche ai punti 1, 2, 3 e 4 e come queste tecniche sono state utilizzate per risolvere problemi nell'ambito "human data" 6. Acquisire la sensibilità necessaria per saper discernere e utilizzare i modelli ai punti 1, 2, 3, e 4 7. Essere in grado di risolvere un problema di analisi dei dati e: 1) sintetizzarne la soluzione in un documento (project report) scritto in modo professionale, 2) presentare il lavoro svolto oralmente nella forma di "presentazione stile conferenza" con dimostrazione del software scritto allo scopo 8. Essere in grado di utilizzare e implementare al computer gli algoritmi ai punti 1, 2, 3 e 4 tramite il linguaggio Python

Attività di apprendimento previste e metodologie di insegnamento:

Il corso è articolato come segue: - Didattica frontale (36 ore): il docente presenterà gli argomenti presenti nel syllabus delle lezioni utilizzando principalmente delle slide. Particolari passaggi matematici verranno presentati tramite l'ausilio della lavagna. --- Una lezione verrà dedicata alla presentazione del progetto del corso, del set di dati da utilizzare e delle modalità di esame --- Il corso è progettato, per ogni sezione, in modo da presentare prima una trattazione teorica dei metodi e dei modelli, per poi descrivere alcune applicazioni di riferimento, spiegando nel dettaglio come i metodi vengono usati / adattati nei vari ambiti, commentando scelte architettoniche e le relative prestazioni - Attività di laboratorio (12 ore): sei lezioni verranno dedicate all'attività di laboratorio, nella quale le tecniche sviluppate durante le lezioni teoriche verranno implementate e caratterizzate sperimentalmente utilizzando il linguaggio Python. Tutto il materiale didattico presentato / utilizzato (slides e software) sarà reso disponibile sul sito Moodle del corso.

Contenuti:

Parte I – Introduzione (2 ore) - Introduzione al corso, modalità di esame, orari del docente, etc. - Applicazioni: salute, servizi che includono il tracciamento di attività, applicazioni di sicurezza e emergenza, analisi di movimento, natural language processing Part II – Tecniche di quantizzazione vettoriale (18 ore) - Quantizzazione Vettoriale (QV): -- Significato, utilità, metriche -- Clustering 1: K-means, soft K-means, Expectation Maximization (EM) -- Clustering 2: X-means, DBSCAN (density based clustering) - Algoritmi per la QV non-supervisionata: --- Mappe Auto-Organizzanti (SOM), Reti Neurali "Gas" (GNG): teoria e algoritmi - Applicazioni a segnali biometrici quasi-periodici (ECG): -- Pre-processamento del segnale, normalizzazione, segmentazione --- Apprendimento di dizionari: concetti, architetture --- Rappresentazione efficiente di segnali ECG: descrizione di tecniche dello stato dell'arte --- Esempi di architetture per l'apprendimento non-supervisionato GNG-based per segnali ECG --- Architettura finale del sistema e risultati numerici Part III - Reti Neurali "Deep" (16 ore) - Review: tecniche di discesa del gradiente e concetti generali (apprendimento supervisionato, modelli di costo, etc.) - Review: Reti neurali "feed forward": modelli, allenamento delle reti, algoritmo di "back-propagation" - review - Reti neurali convoluzionali (CNN): struttura, blocchi costituenti,

allenamento, batch normalization - Architetture autoencoder per l'estrazione automatica di feature -- denoising autoencoder - Reti ricorrenti (RNN) per l'analisi di sequenze temporali. Esempio: natural language processing -- Autoencoder per sequenze temporali basati su RNN - Residual convolutional neural networks (ResNets) - Meccanismi di attenzione per reti CNN (immagini) e RNN (sequenze temporali) - Architetture Transformer: spiegazione passo passo dei principi di funzionamento, dei blocchi architetturali e dei meccanismi per il loro training. Esempi applicativi e trend Part IV: Lezioni di laboratorio (12 ore) Le lezioni di laboratorio saranno effettuate in modo guidato, presentando allo studente codice pre-scritto e funzionante, ma allo stesso tempo soffermandosi su ogni blocco del codice, caratterizzandone il funzionamento e i dati che esso ritorna in uscita. Le lezioni si incentreranno sulla comprensione del linguaggio Python e dei vari tools (Keras e TensorFlow) per la definizione e l'allenamento di reti neurali. Le reti neurali "feed forward", quelle convoluzionali e ricorrenti, verranno pienamente caratterizzate descrivendo i blocchi che le compongono, la loro programmazione e la creazione delle varie architetture neurali. Queste verranno infine allenate tramite diversi algoritmi che sfruttano la discesa del gradiente. Le reti trattate verranno poi testate con opportuni dataset. In estrema sintesi, gli argomenti trattati sono: - Introduzione alla programmazione con il linguaggio Python - Reti neurali con struttura "feed forward" - Reti neurali convoluzionali (CNN) - Reti neurali ricorrenti (RNN) - Meccanismi avanzati (implementazione): batch normalization, residual networks (ResNets), autoencoders, inception layers

Modalità di esame:

Questo è un corso di "machine learning" avanzato e applicato a problematiche reali. Pertanto, la verifica finale dell'apprendimento dello studente verterà su un progetto, che coinvolgerà le seguenti fasi di lavoro: 1. Il Professore assegnerà agli studenti un problema da risolvere, individuando un dataset aperto, pubblicamente fruibile e utilizzabile allo scopo. Il problema verrà descritto dal docente tramite un'apposita lezione, nella quale si spiegheranno altresì le modalità di presentazione dei risultati finali, via 1) una relazione finale, in forma scritta, 2) una presentazione, in forma orale. 2. Gli studenti si distribuiranno in gruppi, con un massimo di due persone per gruppo di lavoro, e inizieranno a lavorare al progetto assegnato. La scelta della tecnica da utilizzare, il pre-trattamento dei dati per ottenere delle feature informative, etc. saranno tutti dettagli da risolvere, a discrezione dello studente. Il docente si renderà disponibile per seguire i gruppi nelle varie fasi di lavoro. 3. Ogni gruppo risolverà il problema proposto secondo la tecnica prescelta e presenterà: 1) una relazione finale, 2) una presentazione al docente in forma orale dove si descriverà, il problema affrontato, i modelli e tecniche utilizzati, il codice realizzato allo scopo, i risultati ottenuti. È inoltre apprezzata la dimostrazione del funzionamento del codice realizzato da parte degli studenti. Il voto finale verrà proposto dal docente dopo un'attenta valutazione della relazione scritta al punto 1) e della presentazione in forma orale al punto 2).

Criteri di valutazione:

I criteri di valutazione con cui verrà effettuata la verifica delle conoscenze e delle abilità attese, saranno: 1. Completezza delle conoscenze acquisite. 2. Capacità di analisi di un problema reale attraverso le tecniche presentate nel corso. 3. Proprietà nella terminologia tecnica usata, sia scritta che orale. 4. Originalità e indipendenza nella identificazione della soluzione scelta per la soluzione del progetto. 5. Competenza e coerenza nell'interpretazione del significato dei risultati ottenuti / qualità degli stessi. 6. Abilità nell'utilizzo degli strumenti informatici nello studio del problema assegnato. 7. Qualità dell'esposizione orale. 8. Qualità dell'esposizione scritta.

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

Report tecnici, articoli scientifici, software (Python code) e tutto il materiale necessario per un corretto apprendimento dei concetti esposti saranno resi disponibili dal docente tramite la pagina Moodle del corso. - Le slide e le registrazioni delle lezioni verranno rese disponibili tramite la pagina Moodle del corso. - Altri libri utili (non obbligatori) sono: Per un ripasso di concetti di algebra lineare: - J. R. Magnus and H. Neudecker, "Matrix Differential Calculus with Applications in Statistics and Econometrics," Wiley, 1999. Per un ripasso dei concetti matematici propedeutici al machine learning: - Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong, "Mathematics for Machine Learning," Cambridge University Press, 2020.

NATURAL LANGUAGE PROCESSING

Titolare: Prof. GIOVANNI DA SAN MARTINO

Mutuato da: Laurea magistrale in Computer Science (Ord. 2021)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Sono richieste conoscenze di base di Analisi Matematica, Probabilità e Statistica, Algebra Lineare. È consigliata familiarità con la programmazione, nello specifico il linguaggio Python. È altresì consigliata familiarità con i concetti di base di Apprendimento Automatico.

Conoscenze e abilità da acquisire:

ABILITÀ COGNITIVE Gli studenti saranno in grado di: - comprendere ed utilizzare una terminologia appropriata - scegliere le tecniche più adatte per risolvere problemi di apprendimento automatico riguardanti il linguaggio naturale - analizzare in modo competente i risultati di propri esperimenti ed articoli in letteratura. **ABILITÀ PRATICHE** Gli studenti saranno in grado di: - utilizzare librerie di apprendimento automatico e natural language processing per risolvere problemi di complessità crescente - utilizzare in modo critico gli strumenti collegando le varie azioni ai concetti teorici corrispondenti. **COMPETENZE TRASVERSALI** Gli studenti svilupperanno capacità di pensiero critico, capacità di inquadrare un problema e di problem solving.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso offrirà: - Lezioni frontali, - Compiti individuali, - esercizi in laboratorio

Contenuti:

Gli argomenti del corso sono i seguenti: - Introduzione agli algoritmi di apprendimento automatico: una prospettiva applicativa; - Elaborazione di base di testo: sentence splitting, tokenisation, Part of Speech tagging, Named Entity Recognition, - Coreference Resolution - Syntactic Parsing, Semantic Role Labeling - Rappresentazione vettoriale di dati testuali: n-grams, word2vec, rappresentazione di frasi. - Generazione di testo: language models, sequence-to-sequence models - Applicazioni: machine translation, sentiment analysis, (community) question answering, news analysis.

Modalità di esame:

Le conoscenze degli studenti e le loro abilità saranno valutate attraverso: - Compiti - gli studenti risolveranno durante il corso dei brevi esercizi sugli argomenti visti a lezione - Un progetto - gli studenti applicheranno le loro conoscenze per risolvere un problema che riguarda il linguaggio naturale - Un esame orale - gli studenti dovranno rispondere a domande sul progetto e/o sugli argomenti visti a lezione

Criteri di valutazione:

Compiti: 25%, Progetto con report scritto: 40%, esame orale: 35%.

Testi di riferimento:

Eisenstein, Jacob, Introduction to natural language processing. Cambridge, Massachusetts London: The MIT Press, 2019 by Daniel Jurafsky; James H. Martin, Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition. : Pearson/Prentice Hall, 2009

Eventuali indicazioni sui materiali di studio:

Verranno resi disponibili i lucidi utilizzati a lezione ed ulteriore materiale per approfondire i concetti visti a lezione.

OPTIMIZATION FOR DATA SCIENCE

Titolare: Prof. FRANCESCO RINALDI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Conoscenze di base in - Analisi reale e Calcolo; - Algebra lineare; - Teoria della probabilità.

Conoscenze e abilità da acquisire:

Comprendere modelli e metodi di ottimizzazione nell'ambito del Data Science. Nello specifico: 1) Comprendere le proprietà teoriche che sono d'interesse per lo sviluppo di modelli matematici in data science. 2) Analizzare e utilizzare modelli matematici esistenti per la risoluzione di problemi reali nell'ambito del data science. 3) sviluppare e/o utilizzare metodi di risoluzione appropriati.

Attività di apprendimento previste e metodologie di insegnamento:

- Il corso si baserà su lezioni frontali; - Il docente utilizzerà la lavagna e le slides; - Dispense e slide verranno rese disponibili sulla piattaforma moodle.

Contenuti:

1. Insiemi convessi e funzioni convesse (a) Convessità, nozioni di base; (b) Funzioni convesse: nozioni di base e proprietà; 2. Ottimizzazione convessa non vincolata: (a) Modelli in Data Science; (b) Caratterizzazione delle soluzioni ottime; (c) Metodi tipo gradiente; (d) Metodi tipo gradiente a blocchi; (e) Metodi per l'ottimizzazione stocastica. 3. Ottimizzazione convessa vincolata; (a) Modelli in Data Science; (b) Caratterizzazione delle soluzioni ottime; (c) Metodi basati su approssimazioni poliedrali; (d) Metodi di proiezione; 4. Ottimizzazione su reti di grande dimensione (a) Modelli di reti in Data Science; (b) Metodi di ottimizzazione distribuita.

Modalità di esame:

- Prova scritta alla fine del corso - Esercizi - Progetto (Opzionale) 1) Gli Esercizi verranno assegnati periodicamente. Gli studenti dovranno consegnare le soluzioni a scadenze prestabilite. 2) La prova d'esame è scritta e prevede 4 domande a risposta aperta. 3) Il progetto (opzionale) può essere richiesto per approfondire tematiche specifiche. La prova scritta contribuisce all' 85% del voto. Gli esercizi contribuiscono al 15% del voto. Il progetto può integrare o sostituire lo scritto.

Criteri di valutazione:

La valutazione della preparazione dello studente si baserà: - sulla comprensione degli argomenti svolti in aula; - sull'acquisizione dei concetti di carattere teorico; - sulla capacità di utilizzare in maniera autonoma e consapevole i modelli e le metodologie risolutive proposte.

Testi di riferimento:

Nesterov, Yurii, Introductory lectures on convex optimization: A basic course.. : Vol. 87. Springer Science & Business Media, 2013

Eventuali indicazioni sui materiali di studio:

- Dispense e slide fornite dal docente.

PROCESS MINING

Titolare: Prof. MASSIMILIANO DE LEONI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Conoscenze di base di algoritmi, strutture dati e programmazione, come acquisiti nel corso "Fundamental of Information Systems" o nei corsi di laurea in Informatica o affini.

Conoscenze e abilità da acquisire:

Oggigiorno, tutte le medie e grandi imprese seguono procedure e processi standardizzati per fornire beni e servizi, al fine di garantire livelli di servizio e di efficienza più alti. La modellazione formale di questi processi viene utilizzata come input per sistemi informativi (per es. sistemi SAP o ERP) che automaticamente si autoconfigurano per funzionare in accordo con tali modelli di processo. Lo scopo del corso è quello di introdurre e studiare concetti, metodi e tecniche che riguardano la modellazione, l'analisi e il miglioramento di processi di business. Processi più efficienti portano a sistemi più efficienti, che a loro volta permettono alle organizzazioni di fornire migliori servizi e beni di più alta qualità. Il corso può concettualmente essere suddiviso in tre parti: 1) In una prima parte, il corso introduce i principi e i linguaggi alla base della modellazione dei processi. 2) In una seconda parte, la più estesa, il corso discute le tecniche di analisi dei processi aziendali basate sul Process Mining. Process Mining si pone l'obiettivo di utilizzare lo storico dei dati transazionali delle aziende (i log degli eventi) per analizzare e migliorare i processi aziendali. Il Process Mining utilizza le informazioni nei log degli eventi per scoprire come i processi vengono REALMENTE eseguiti, svelando le violazioni di protocolli e regolamenti, le inefficienze e tutte le criticità che mettono a rischio le esecuzioni stesse dei processi. Mentre le tecniche tradizionali di Business Intelligence si basano su interviste e questionari con gli attori dei processi, Process Mining è parte integrante di Data Science: la soggettività degli attori viene sostituita dall'oggettività delle conclusioni estratte dai dati effettivi

di processo. 3) L'obiettivo ultimo del Process Mining è predisporre azioni concrete che possano portare a migliorare i processi, sia dal punto di vista di aderenza alle normative vigenti, sia dal punto di vista dell'efficienza. Nella terza parte del corso, gli "insight" ottenuti tramite Process Mining vengono utilizzati per costruire modelli di simulazioni dei processi aziendali che siano "gemelli digitali" dei processi reali. Tali "gemelli digitali" rispecchiano il comportamento dei processi reali, in tutti gli aspetti (caratteristiche e vincoli temporali, decisioni, dati manipolati, risorse, ecc.). Il vantaggio di avere modelli di simulazione (i gemelli digitali) risiede nel fatto che sia è possibile simulare variazioni di tali processi e misurare gli impatti sulle performance dei processi, senza dover veramente cambiare il processo reale, riducendo i costi e rischi di eseguire processi alternativi ancora più costosi e meno efficienti degli originali.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso consiste di lezioni, esercitazioni, sessioni di laboratorio e seminari. Le sessioni di laboratorio daranno l'opportunità di "sporcarsi le mani" e verranno prevalentemente utilizzati dataset reali per meglio comprendere la complessità di eseguire un progetto di Process Mining e di lavorare con dati di processo reali. I seminari verranno organizzati da aziende nazionali e internazionali del settore di Process Mining

Contenuti:

Il corso affronterà i seguenti argomenti: 1. MODELLAZIONE DEI PROCESSI TRAMITE RETI DI PETRI - Concetti di base di reti di Petri - Uso delle reti di petri per la modellazione dei processi aziendali - Analisi strutturale di reti di Petri - Correttezza di modelli di processi aziendali basati su reti di Petri: Principi e Verifica. - Uso di Woped come software per modellare e verificare la correttezza dei modelli di processo 2. PROCESS MINING - Introduzione su Process Mining e il log degli eventi - Tecniche di Base per Process Discovery e loro Limiti - Tecniche Avanzate per Process Discovery: Heuristic Miner and Region Miner - Verifica di Conformità basata su Token Replay - Verifica di Conformità basata sugli Allineamenti - Il Mining delle Prospettive Aggiuntive su Decisioni, Tempistiche e Risorse - Analisi delle Reti Social - Uso di software open-source and commerciali per l'analisi dei processi aziendali tramite Process Mining 3. SIMULAZIONE DEI PROCESSI AZIENDALI - Principi e Metodologie per la simulazione: Definizione di intervalli di Warm-up e Cool-down, gestione della stocasticità della simulazione tramite simulazioni multiple ed indipendenti, analisi dei risultati. - Definizione degli elementi di uno scenario di simulazione di processi aziendali: arrival rate, branching probabilities, resource pools, distribuzioni di probabilità delle durate delle attività. - Uso di BPMN per la specifica di modelli di simulazione di processi aziendali - Uso di Process Mining per la generazione di modelli di simulazione - Metodologie per la definizione scenari "What-if" - Uso di BIMP come software per la simulazione di processi aziendali.

Modalità di esame:

Esame scritto e Progetto

Criteri di valutazione:

Il lavoro di progetto e l'esame scritto saranno valutati sulla base dei seguenti criteri: i) conoscenza dello/a studente/ssa di concetti, metodi e tecnologie; ii) abilità dello/a studente/ssa di padroneggiare le tecnologie di implementazione; iii) le capacità di sintesi, precisione, chiarezza e astrazione dello/a studente/ssa dimostrate nell'esame scritto e nella presentazione di progetto. Il voto finale è ottenuto come la somma pesata dei voti dell'esame scritto (60%) e del progetto (40%).

Testi di riferimento:

Marlon Dumas, Marcello La Rosa, Jan Mendling, Hajo A. Reijers, Fundamentals of business process management. Berlin: Springer, 2018 Aalst, Wil M. P. : van der, Modeling business processes a Petri net-oriented approach. Cambridge: MIT Press, 2011 van der Aalst, Wil, Process Mining Data Science in Action. Berlin, Heidelberg: Springer Berlin Heidelberg Imprint: Springer, 2016

Eventuali indicazioni sui materiali di studio:

Tre testi verranno utilizzati per le tre parti, come indicato nel box "Testi di riferimento" sottostante Questi sono complementati da (1) un eserciziario e da (2) documenti relativi ad esercitazioni autoguidate per la parte di laboratorio. Sia (1) che (2) saranno messi a disposizione su Moodle.

SEMINARS AND OTHER ACTIVITIES

Titolare: Prof. WOLFGANG ERB

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, annuale

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: ; 3,00

STATISTICAL LEARNING (C.I.)

Titolare: Prof. ALBERTO ROVERATO

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Prerequisiti:

basic probability theory; multivariable calculus; linear algebra; basic computing skills

Conoscenze e abilità da acquisire:

become familiar with statistical thinking; gain adequate proficiency in the development and use of standard statistical inference tools; be able to analyse datasets using a modern programming language such as R

Modalità di esame:

written test and project work

Criteri di valutazione:

the successful student should show knowledge of the key concepts, skills in the analysis of data and competency in applications

Moduli del C.I.:

Statistical Learning 1 (Mod. A)
Statistical Learning 2 (Mod. B)

STATISTICAL LEARNING 1 (MOD. A)

Titolare: Prof. ALBERTO ROVERATO

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 48A; 6,00

STATISTICAL LEARNING 2 (MOD. B)

Titolare: Prof. ALBERTO ROVERATO

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 48A; 6,00

STOCHASTIC METHODS

Titolare: Prof. MARCO FERRANTE

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Nozioni di base di calcolo differenziale e integrale, algebra lineare e calcolo delle probabilità.

Conoscenze e abilità da acquisire:

Lo scopo del corso è di introdurre metodi e concetti di Calcolo delle Probabilità e di alcune classi di Processi Stocastici che hanno un forte impatto come strumenti algoritmici, computazionali e nello studio delle reti.

Attività di apprendimento previste e metodologie di insegnamento:

Lezioni frontali integrate da materiale online. Esercitazioni in aula

Contenuti:

1. Richiami di Calcolo delle Probabilità. • spazi di probabilità, sigma-algebre, proprietà della probabilità • probabilità condizionata e indipendenza • variabili aleatorie, valor atteso e valor atteso condizionale • Legge dei grandi numeri e Chernoff Bounds • approssimazione di distribuzioni di probabilità. 2. Catene di Markov e passeggiate aleatorie • Catene di Markov a tempo discreto: definizione e classificazione degli stati • distribuzioni stazionarie e risultati di convergenza. 3. Martingale a tempo discreto • Proprietà del valore atteso condizionato • Definizione di martingala, sopra- e sotto martingala. • Tempi di arresto

Modalità di esame:

Esame scritto

Criteri di valutazione:

Il voto finale è basato sul risultato della prova scritta, il cui scopo principale è verificare la capacità di usare in modo corretto ed efficiente le tecniche esposte, applicandole a problemi concreti.

Testi di riferimento:

Paolo Dai Pra, Stochastic Methods for Data Science. ,

Eventuali indicazioni sui materiali di studio:

Il docente fornirà delle dispense ed esercizi

Curriculum: Data Science - Piano di studio Machine Learning for Intelligent systems**ADVANCED TOPICS IN DATA SCIENCE**

Titolare: Prof. LAMBERTO BALLAN

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, annuale

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 24A; 3,00

BIG DATA COMPUTING

Titolare: Prof. FRANCESCO SILVESTRI

Mutuato da: Scuola Galileiana di Studi Superiori - Classe di Scienze Naturali

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Il corso ha i seguenti prerequisiti: buone competenze relative al progetto e all'analisi di algoritmi e strutture dati; conoscenza delle nozioni fondamentali di calcolo delle probabilità e statistica; buon livello di programmazione in Java o Python; capacità di utilizzo dell'ambiente Linux e della linea di comando.

Conoscenze e abilità da acquisire:

In questo corso gli studenti imparano tecniche algoritmiche fondamentali per l'elaborazione efficiente ed efficace di insiemi di dati di grande dimensione. Inoltre, attraverso alcune attività pratiche, essi acquisiscono abilità relative allo sviluppo di applicazioni in Apache Spark, che è uno dei framework di programmazione più popolari e diffusi per big data computing.

Attività di apprendimento previste e metodologie di insegnamento:

Lezioni frontali, uso di piattaforme di students engagement, seminari di esperti selezionati, e attività propedeutiche allo svolgimento degli homework.

Contenuti:

Il corso affronterà i seguenti argomenti: Introduction to the Big Data phenomenon. * Frameworks: - Distributed (MapReduce, Apache Spark) - Streaming * Techniques with applications: - Partitioning (data distribution) - Coresets (unsupervised learning) - Sketches (estimation of moments, set membership) - Locality sensitive hashing (similarity search).

Modalità di esame:

L'esame consiste in alcuni homework di programmazione, assegnati ogni 2-3 settimane e da svolgere in gruppi di 2-3 studenti, e in una prova scritta individuale comprendente domande teoriche ed esercizi.

Criteri di valutazione:

La valutazione finale è basata sugli homework e sulla prova scritta. Gli homework mirano a verificare la capacità degli studenti di programmare applicazioni big data in Apache Spark, mentre la prova scritta mira a verificare la loro conoscenza delle tecniche algoritmiche apprese durante il corso e la loro capacità di problem solving nel contesto big data.

Testi di riferimento:

Leskovec, Jure; Rajaraman, Anand; Ullman, Jeffrey D.; Leskovec, Jure, Mining of massive datasets. Cambridge: Cambridge University Press, 2020

Eventuali indicazioni sui materiali di studio:

Il diario delle lezioni, il materiale didattico e le modalità d'esame dettagliate sono resi disponibili sul MOODLE del corso e sul MOODLE esami.

BIOINFORMATICS

Titolare: Prof.ssa LAURA TREU

Mutuato da: Laurea magistrale in Computer Science (Ord. 2021)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 40A+8E; 6,00

Prerequisiti:

Non ci sono prerequisiti particolari, se non quanto ci si aspetta da uno studente magistrale di informatica. Una conoscenza di base della genetica e della biologia molecolare saranno comunque utili per meglio inquadrare le motivazioni biologiche che stanno alla base della bioinformatica. Il corso è in lingua inglese, quindi è necessario avere una buona conoscenza dell'inglese scritto e parlato.

Conoscenze e abilità da acquisire:

Il Corso è suddiviso in tre parti principali: la prima parte mette in relazione Biologia e Informazione; la seconda parte descrive i principali algoritmi utilizzati in bioinformatica per allineare sequenze biologiche e assemblare genomi; la terza parte tratta di problemi di bioinformatica relativi alla genomica funzionale. Inoltre il corso è accompagnato da esercitazioni pratiche in cui gli studenti applicheranno metodi bioinformatici per analizzare dati genomici. In considerazione della complessità della materia e in accordo con i descrittori di Dublino, particolare attenzione sarà dedicata affinché gli studenti acquisiscano la capacità di integrare le conoscenze e gestire la complessità dei problemi trattati, nonché di formulare giudizi sulla base di informazioni limitate e spesso frammentarie.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso sarà tenuto con lezioni frontali e con esercitazioni pratiche. L'attività didattica sarà supportata da risorse messe a disposizione sulla piattaforma e-learning "Moodle", che comprendono materiale per apprendimento remoto e per auto-valutazione. In questo modo si vuole promuovere un'attività di "blended learning" con cui lo studente, almeno in parte, impara autonomamente, seguendo un percorso che lo accompagna attraverso contenuti reperibili

in rete. Dove possibile si applicherà il paradigma della "flipped classroom" che inverte lo schema tradizionale di insegnamento, prevedendo che prima lo studente impari la lezione autonomamente per poi discutere e approfondire gli argomenti in classe, con il docente e con gli altri studenti. Un'ampia raccolta di problemi, questionari ed esercizi viene messa a disposizione sulla piattaforma Moodle, sia per consentire l'autovalutazione, sia per stimolare argomenti di discussione da approfondire in classe.

Contenuti:

Questo è un corso di 6 crediti: cinque di lezioni ed uno di attività pratiche che consistono nell'implementazione di algoritmi oppure in un'approfondita indagine della letteratura su argomenti assegnati. Le lezioni sono organizzate in tre parti. La prima parte è un'approfondita introduzione alla Biologia, presentata come una disciplina scientifica centrata sull'Informazione. I meccanismi che facilitano la trasmissione e l'evoluzione dell'informazione biologica saranno presi come spunto per introdurre alcuni problemi della biologia che richiedono approcci computazionali e strumenti bioinformatici. La seconda parte del corso descrive i principali algoritmi utilizzati per allineare sequenze biologiche, inclusi quelli sviluppati per il sequenziamento di DNA di ultima generazione. Sono inoltre descritti gli algoritmi utilizzati per l'assemblaggio "de novo" di genomi. Infine, la terza parte del corso copre alcuni aspetti della bioinformatica relativi alla genomica funzionale, come la predizione e annotazione genica, l'analisi del trascrittoma e la loro integrazione attraverso modelli computazionali del metabolismo. Inoltre viene discusso il ruolo della bioinformatica nell'analisi di genomi individuali e nella medicina personalizzata.

Modalità di esame:

L'esame si svolgerà in modalità scritta.

Criteri di valutazione:

Nell'esame finale gli studenti dovranno dimostrare una comprensione sistematica del settore e dovranno sapersi destreggiare con i metodi della ricerca associati ad esso. Inoltre gli studenti dovranno essere capaci di analisi critica, di valutare e sintetizzare idee nuove e complesse, integrando gli argomenti di questo corso con altre conoscenze.

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

Non sono previsti libri ufficiali di testo e gli studenti saranno stimolati a trovare le informazioni su fonti multiple. Il materiale didattico sarà messo a disposizione sulla piattaforma e-learning Moodle.

BIOLOGICAL DATA

Titolare: Prof. DAMIANO PIOVESAN

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Conoscenze base di informatica, algoritmi di ottimizzazione e machine learning. Linguaggio di programmazione: Python.

Conoscenze e abilità da acquisire:

Il corso intende comunicare le conoscenze di base sui diversi tipi di dati biologici, come sequenze, strutture, reti e letteratura. Inoltre intende permettere allo studente lo svolgimento autonomo di un progetto di ricerca in bioinformatica, definendo lo stato dell'arte per un problema aperto e un tentativo di risolverlo con lo sviluppo di software che estenda librerie esistenti e la valutazione critica dei risultati ottenuti.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso si compone di lezioni frontali, esercitazioni pratiche al computer, sviluppo di un progetto e presentazione dello stesso con discussione critica. Le esercitazioni servono per familiarizzare lo studente con le librerie software da usare per un progetto bioinformatico relativo ad un problema attuale diverso per ogni gruppo. La presentazione del progetto richiede una discussione in cui far emergere i punti di forza e debolezza della soluzione prescelta.

Contenuti:

Il corso si compone di diverse parti, che corrispondono ai diversi tipi di dati biologici: Sequenze - DNA e proteine - Allineamenti - Ricerca veloce in database Strutture - Proteine non globulari - Motifs discovery - Predizione di disordine Funzione - Knowledgebases - Predizione di funzione Letteratura - Artticoli scientifici - Text mining

Modalità di esame:

L'esame è suddiviso in quattro parti separate, ognuna delle quali deve essere superata per completare con successo il corso. I pesi relativi di ciascuna parte sono indicati tra parentesi di seguito: 1. Test scritto a metà corso, che consiste in domande teoriche e pratiche (20%) 2. Journal Club, presentazione di un articolo scientifico inerente al corso (20%) 3. Progetto software (30%) 4. Presentazione del progetto e valutazione critica (30%) Per superare l'esame, gli studenti devono completare con successo tutte e quattro le parti e ottenere un voto sufficiente in ciascuna di esse. Il test scritto a metà corso valuterà la comprensione degli studenti dei concetti teorici affrontati nel corso, mentre il progetto software valuterà la loro capacità di applicare questi concetti a un problema pratico. La presentazione del progetto valuterà invece la capacità degli studenti di comunicare i risultati della loro ricerca e di valutare il lavoro dei loro colleghi.

Criteri di valutazione:

- Comprensione di concetti e gli algoritmi presentati a lezione - Capacità di applicare le nozioni fornite a lezione su problemi reali - Capacità critica di saper utilizzare i metodi nei modi più opportuni, scegliendo tra le alternative possibili - Capacità di sviluppare software riutilizzabile estendendo librerie esistenti - Capacità espositiva e di discussione critica

Testi di riferimento:

Krogh, Anders; Eddy, Sean R.; Mitchison, Graeme; Durbin, Richard; Krogh, Anders, Biological sequence analysis probabilistic models of proteins and nucleic acids. : , 1998

Eventuali indicazioni sui materiali di studio:

Sul sito E-learning vengono resi disponibili molti materiali per il corso. Questi comprendono i lucidi del corso (appena disponibili) e le registrazioni audio (podcast), le dispense e la letteratura usata per i progetti. Le lezioni pratiche sono organizzate tramite CoLab notebook in Python.

BUSINESS ECONOMIC AND FINANCIAL DATA

Titolare: Prof.ssa MARIANGELA GUIDOLIN

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

COGNITION AND COMPUTATION

Titolare: Prof. MARCO ZORZI

Mutuato da: Laurea magistrale in Cybersecurity (Ord. 2020)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Il corso richiede conoscenze preliminari sull'apprendimento automatico e sulla teoria della probabilità. La familiarità con concetti base di psicologia cognitiva e neuroscienze può facilitare la comprensione dei temi trattati.

Conoscenze e abilità da acquisire:

Il corso fornisce conoscenze sui principali approcci computazionali utilizzati per modellizzare l'apprendimento e la cognizione umana, dalle reti neurali artificiali ai modelli probabilistici strutturati. Queste conoscenze sono rilevanti sia per la comprensione del funzionamento della mente che per il disegno di moderni sistemi di intelligenza artificiale. La discussione teorica dei diversi approcci verrà affiancata da esempi concreti di applicazione a problemi di modellizzazione cognitiva.

Attività di apprendimento previste e metodologie di insegnamento:

L'insegnamento è basato su lezioni frontali in cui vengono trattati gli argomenti teorici. Verranno utilizzate tecniche di apprendimento interattivo, come le discussioni interattive su domande aperte, per promuovere la riflessione critica sui concetti discussi. Il corso include alcune esercitazioni (in aula informatica) con simulazioni al computer.

Contenuti:

1. Introduzione: modellizzazione computazionale e matematica nelle scienze cognitive e nelle neuroscienze cognitive. Rassegna degli approcci simbolici, emergentisti e probabilistici alla simulazione della cognizione umana. 2. Modelli probabilistici della cognizione: concetti base su inferenza Bayesiana e modelli grafici probabilistici; apprendimento induttivo; programmazione probabilistica. 3. Modelli connessionisti della cognizione: concetti base sulla computazione neurale; apprendimento nelle reti neurali; architetture per deep learning. 4. Codifica dell'informazione nelle architetture cognitive: codifica efficiente, codifica probabilistica, codifica predittiva. 5. Casi di studio: modelli della percezione umana e dell'acquisizione di concetti; acquisizione del linguaggio e comprensione del linguaggio naturale; ragionamento causale e decision making.

Modalità di esame:

L'esame consisterà in una prova scritta con domande aperte e domande a scelta multipla. Ogni studente dovrà preparare inoltre un elaborato scritto individuale (o report di progetto) assegnato durante il corso e consegnarlo (tramite il Moodle del corso) entro il giorno precedente l'esame scritto.

Criteri di valutazione:

La valutazione sarà basata sulla comprensione degli argomenti trattati nel corso e sull'acquisizione dei concetti e delle metodologie proposte.

Testi di riferimento:

Koller, Daphne; Friedman, Nir, Probabilistic Graphical Models. Cambridge: MIT Press, 2009 Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron, Deep learning. London, England; Cambridge, Mass: The MIT Press, 2016 Russell, Stuart; Norvig, Peter, Artificial Intelligence. Harlow: Pearson Education UK, 2013

Eventuali indicazioni sui materiali di studio:

Tutti gli argomenti verranno trattati durante le lezioni. Le slides delle lezioni saranno disponibili sulla piattaforma di e-learning Moodle. Gli appunti degli studenti dovranno essere integrati dai libri di testo e da altro materiale (articoli scientifici) forniti dal docente sulla piattaforma di e-learning.

DEEP LEARNING (MATICOLE DISPARI)

Titolare: Prof. ALESSANDRO SPERDUTI

Mutuato da: Laurea magistrale in Computer Science (Ord. 2021)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

DEEP LEARNING (MATICOLE PARI)

Titolare: Prof. ALESSANDRO SPERDUTI

Mutuato da: Laurea magistrale in Computer Science (Ord. 2021)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

ENGLISH LANGUAGE B2 (PRODUCTIVE SKILLS)

Titolare: Prof. WOLFGANG ERB

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: II anno, annuale

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: ; 3,00

FUNDAMENTALS OF INFORMATION SYSTEMS

Titolare: Prof. GIORGIO MARIA DI NUNZIO

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 96A; 12,00

HIGH DIMENSIONAL PROBABILITY FOR DATA SCIENCE

Titolare: Prof. ALBERTO CHIARINI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

INFORMATION RETRIEVAL

Titolare: Prof. NICOLA FERRO

Mutuato da: Scuola Galileiana di Studi Superiori - Classe di Scienze Naturali

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Requested competencies: - good and proactive programming skills and, in particular, the object-oriented paradigm and its design principles; - good knowledge of the Java programming language;

Conoscenze e abilità da acquisire:

The objective of the course is to learn the methodologies for the design, development, and evaluation of information retrieval systems, search engines, and recommender systems, practicing them through the development of an actual application, e.g. for searching and recommending newspaper articles. This objective calls for acquiring: + strong and advanced computer science competencies on information retrieval, design and development methodologies, architectural alternatives, and performance evaluation techniques; + good knowledge of the features and functioning of information retrieval systems, search engines, recommender systems and, more in general, information access systems; + capability of developing a real information retrieval application, using the Java language, and of evaluating its performance, also through statistical techniques.

Attività di apprendimento previste e metodologie di insegnamento:

+ Lectures + Labs + Seminars of visiting colleagues on research topics and/or seminar by companies on the use and perspectives for innovative products based on information retrieval systems, stage opportunities, simulation of job interviews. + Homeworks: to design and develop a "real" information retrieval application. Homework deadlines are aligned with schedule and the contents of the lectures so that students can immediately apply, during the course, the learned concepts to a case study of their own interests. + Oral presentation with slides and demo of the homework project

Contenuti:

The course covers the full spectrum of information retrieval methodologies, starting from foundational approaches to offline system development, continuing with the new techniques for machine learning for search, and concluding with the most recent advances for online system development. More in detail, the course will address: + Offline system development: the student learns the foundations of information retrieval, covering text processing and indexing for efficient access and retrieval; and, basic and advanced retrieval models to match natural language queries to documents. The student learns how to evaluate the performance of an information retrieval system in terms of effectiveness, how to tune the different system parameters, and how to statistically validate the experimental results. Finally, the student learns the principles of Web search engines, Web crawling, search engine optimization, and knowledge graphs and semantic search. + Machine learning for search: the student learns the advanced applications of machine learning techniques for information search and filtering, namely recommender systems, learning-to-rank, and neural information retrieval. + Online system development: the student learns the principles of retrieval as interaction, i.e. the basics of search as a learning process for man and machine, by leveraging click models, A/B testing, interleaved comparison, and online learning-to-rank.

Modalità di esame:

Individual oral exam with questions and exercises on the topics covered during the lectures, including the actual coding of a search engine. If a large number of students will be attending the exam, part of the exam may be turned into a Moodle quiz. Project to design, develop, implement, code, document, and evaluate an actual information retrieval application, carried out in student groups via homeworks during the semester: + git repository containing the project source code and all the related material + report on the contents and state of the project as well as the evaluation of the performance of the developed system; + oral presentation of the project outcomes by using slides; + demo of the developed application.

Criteri di valutazione:

The evaluation will be based on the comprehension and knowledge of the notions and methodologies specific to information retrieval, on the capability of facing the different phases of the design, development, implementation, and coding of information retrieval applications, on the capability of evaluating the performance of such applications, and on the implementation of a project for the development of an information retrieval application.

Testi di riferimento:

Manning, C. D., Raghavan, P., and Schütze, H., Introduction to Information Retrieval. Cambridge, UK: Cambridge University Press, 2008 Croft, W. Bruce; Metzler, Donald; Strohman, Trevor, Search engines: Information Retrieval in Practice. Boston, USA: Pearson, 2009

Eventuali indicazioni sui materiali di studio:

The teaching material consists of: - reference book - instructor's slides - suggested readings - additional material shared during lectures All the teaching material will be available on the Moodle platform (<https://elearning.dei.unipd.it/>) of the course. Examples of teaching material and videos from the course are available at: <https://iiaa.dei.unipd.it/education/search-engines/> Suggested readings: + Büttcher, S., Clarke, C. L. A., and Cormack, G. V. (2010). Information Retrieval: Implementing and Evaluating Search Engines. The MIT Press, Cambridge (MA), USA. + Manning, C. D., Raghavan, P., and Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press, Cambridge, UK.

INTERNSHIP

Titolare: Prof. TOMASO ERSEGHE

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, annuale

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: ; 18,00

Contenuti:

Durante il corso di laurea magistrale in Data Science, gli studenti svolgono uno stage presso un'azienda, un'industria, un centro di ricerca o un laboratorio universitario. Il tirocinio permette di sviluppare nuove competenze, entrare in contatto con l'ambiente di lavoro quotidiano, sviluppare la consapevolezza delle esigenze del mercato e dei clienti. Queste abilità sono molto apprezzate dal mercato del lavoro, rafforzano il curriculum degli studenti e migliorano le loro opportunità di lavoro. Siamo collegati con una serie di selezionati partner industriali e di ricerca, in Italia e nel mondo. Il tirocinio si svolge solitamente durante il 2° anno, 2° semestre e, seppur non obbligatorio, è spesso abbinato alla tesi di laurea magistrale Una breve guida agli stage è disponibile qui: https://datascience.math.unipd.it/files/a-quick-guide-to-internships-@-data-science_final.pdf Una guida dettagliata per gli studenti iscritti è disponibile nella piattaforma elearning: <https://elearning.unipd.it/math/course/view.php?id=659> Per ulteriori richieste contattaci a stage.datascience@math.unipd.it

Testi di riferimento:

CONTENUTO NON PRESENTE

KNOWLEDGE REPRESENTATION AND LEARNING

Titolare: Dott. LUCIANO SERAFINI

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

Prerequisiti:

E' opportuno che gli studenti abbiano una conoscenza di base di logica matematica e di statistica.

Conoscenze e abilità da acquisire:

Abilità cognitive. Al termine del corso, gli studenti saranno in grado di: - Comprendere i principi delle logiche per la rappresentazione e il ragionamento della conoscenza. - Comprendere i principi dell'apprendimento statistico relazionale. - Comprendere i principi dell'integrazione neuro-simbolica, ovvero come combinare logiche e apprendimento statistico relazionale per costruire sistemi di apprendimento e ragionamento in domini ibridi. Abilità pratiche. Gli studenti saranno in grado di: - Modellare problemi in logiche come la Logica Proposizionale, la Logica del Primo Ordine e la Logica Fuzzy. - Utilizzare algoritmi di rappresentazione della conoscenza e di ragionamento per trovare i modelli di una teoria logica. - Modellare problemi utilizzando approcci di apprendimento statistico relazionale come modelli grafici, reti logiche di Markov, Prolog probabilistico e reti logiche tensoriali. Abilità trasversali. Gli studenti svilupperanno capacità di pensiero analitico, capacità di problem solving e capacità di pensiero sistemico-logico.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso prevede 48 ore di lezione. Le lezioni saranno supportate da esercitazioni pratiche.

Contenuti:

(A) Logiche per la rappresentazione della conoscenza: (A.i) introduzione alle logiche proposizionali, sintassi, semantica, procedura decisionale. Soddisfabilità, soddisfabilità ponderata e migliore soddisfabilità. (A.ii) Logica del primo ordine, sintassi, semantica, risoluzione e unificazione. (A.iii) Logiche fuzzy, sintassi, semantica e ragionamento. (B) Apprendimento statistico relazionale: (B.i) Modelli grafici (B.ii) Reti logiche di Markov (B.iii) Prolog probabilistico, (B.iv) Reti logiche tensoriali

Modalità di esame:

Le conoscenze e le competenze degli studenti saranno valutate attraverso: - Un esame scritto: agli studenti verrà chiesto di risolvere esercizi simili a quelli illustrati nelle lezioni. - Un project work: agli studenti verrà chiesto di implementare e valutare uno degli algoritmi presentati nelle lezioni o di utilizzare uno strumento di inferenza logica/probabilistica per risolvere un problema pratico.

Criteri di valutazione:

Gli studenti saranno valutati in base alla conoscenza critica degli argomenti del corso. I criteri di valutazione consistono in: 50% Esame scritto: capacità di presentare e applicare il materiale studiato. 50% Lavoro di progetto: capacità di applicare il contenuto del corso a problemi pratici.

Testi di riferimento:

Russell, Stuart J.; Norvig, Peter; Russell, Stuart J., Artificial intelligence a modern approach. Harlow: Pearson Education, 2022 Stuart Russell, Unifying logic and probability. : Communications of the ACM, 2015

Eventuali indicazioni sui materiali di studio:

Il materiale di studio consiste in: - diapositive delle lezioni, - appunti delle lezioni.

LAW AND DATA

Titolare: Dott.ssa FIORELLA DAL MONTE

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

Prerequisiti:

No prerequisites

Conoscenze e abilità da acquisire:

The course aims to introduce non-law students to a proper understanding of the main legal issues related to the processing of data, personal and non. The first part of the course aims to enable students to approach EU personal data protection regulation. In the second part, instead, students will reflect on the main problems related to the use of data-intensive technologies (big data and artificial intelligence) and the technical and legal solutions now debated.

Attività di apprendimento previste e metodologie di insegnamento:

Classes Seminars Workshops Preassigned readings.

Contenuti:

All the info about the course are on Moodle - Introduction to Law and Legal Studies - Introduction to the EU Law - Introduction to the EU GDPR - The concept of data; personal, sensitive and economic data; big data - Property of data, choices in the management of data - The right to be forgotten - Civil and criminal aspects of profiling activity - Automatic data processing, human responsibilities - The Data Protection Officer and DP Authorities - Civil and criminal protection of privacy - Sanctioning powers and system - Open Data for the public interest - Big data (collection, analysis, processing) and their influence on fundamental rights - Digital Surveillance - Facial Recognition: Open Issues - Disinformation - Artificial Intelligence in the EU law

Modalità di esame:

Written Exam

Criteri di valutazione:

The grading scale used to assess the students is the Italian one, with the highest score of 30/30 and a minimum score of 18/30 (sufficient) (info: here). The students will be graded according to their level of theoretical and practical knowledge of the fields covered throughout the course and their capacity to critically reflect on the most contentious legal issues on data-intensive technologies.

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

The course has no official textbooks. Students can study on their notes and the additional material provided by the instructor. Nevertheless, here are some helpful handbooks to approach some modules. These books are not mandatory, and students have sole discretion to refer them. ? Mireille Hildebrandt (2020). Law for Computer Scientists and Other Folk, OUP (open access: here) – especially chapters 2-3-4-5-9-10 ? Paul Voigt, Axel von dem Bussche (2017). The EU General Data Protection Regulation (GDPR). A practical guide, Springer (unipd access: here) ? European Fundamental Rights Agency (2018). Handbook on European data protection law, Luxemburg (open access: here) ? Karen Yeung, Martin Lodge (2019). Algorithmic Regulation, OUP (Public Law Dept. Library) – especially chapters 2-3-4-6-7-11

MACHINE LEARNING

Titolare: Dott. DAVIDE RIGONI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

NATURAL LANGUAGE PROCESSING

Titolare: Prof. GIOVANNI DA SAN MARTINO

Mutuato da: Laurea magistrale in Computer Science (Ord. 2021)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

OPTIMIZATION FOR DATA SCIENCE

Titolare: Prof. FRANCESCO RINALDI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

PROCESS MINING

Titolare: Prof. MASSIMILIANO DE LEONI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

REINFORCEMENT LEARNING

Titolare: Prof. GIAN ANTONIO SUSTO

Mutuato da: Scuola Galileiana di Studi Superiori - Classe di Scienze Naturali

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

Prerequisiti:

I prerequisiti del corso includono la conoscenza dei fondamenti del machine learning (ad esempio coperti dal corso 'Machine Learning'). Sono inoltre previste conoscenze di elementi di probabilità e statistica, calcolo e algoritmi di ottimizzazione. È necessaria precedente esperienza di programmazione di base con Python per gli homework.

Conoscenze e abilità da acquisire:

Uno studente che completa con successo il corso sarà in grado di definire gli aspetti chiave che differenziano il Reinforcement Learning dagli altri approcci di apprendimento automatico. Dato un problema reale, lo studente sarà in grado di determinare (i) se può essere adeguatamente formulato come un problema di Reinforcement Learning; (ii) essere in grado di formalizzarlo come tale e (iii) identificare un insieme di tecniche più adatte a risolvere il task, insieme agli strumenti software necessari per implementare la soluzione.

Attività di apprendimento previste e metodologie di insegnamento:

Lezioni frontali teoriche, accompagnate da esempi pratici, esercizi numerici ed esercitazioni in Python. Lo studente deve accompagnare allo studio teorico un adeguato numero di esercizi numerici e non. Il percorso dello studente viene guidato da esercitazioni in laboratorio, man mano che i vari argomenti vengono affrontati durante il corso.

Contenuti:

Il corso inizierà con i concetti di base dell'apprendimento in un problema decisionale sequenziale, che saranno formalizzati attraverso i Markov Decision Processes. Verrà esaminata la pianificazione tramite programmazione dinamica, quindi si passerà a trattare predizione e controllo basandosi su approcci di tipo model-free. Saranno trattati gli approcci di approssimazione della value-function, così come i metodi di tipo Policy Gradient. Saranno trattate strategie per l'integrazione di apprendimento e pianificazione e per affrontare il compromesso tra Exploration e Exploitation. Verranno trattati approcci di Imitation Learning e gli approcci di Deep Learning per il Reinforcement Learning (Deep Reinforcement Learning).

Modalità di esame:

La verifica delle conoscenze e delle abilità attese viene effettuata con una prova d'esame scritta che prevede: a) domande aperte sulle tematiche viste

durante le lezioni; b) risoluzioni di esercizi numerici. In aggiunta gli studenti possono sostenere, in via facoltativa, un progetto basato su linguaggio Python.

Criteri di valutazione:

I criteri di valutazione con cui verrà effettuata la verifica delle conoscenze e delle abilità attese, sono: 1. Completezza delle conoscenze acquisite 2. Capacità analitiche nella risoluzione di esercizi 3. Proprietà nella terminologia tecnica usata 4. Qualità dell'esposizione scritta

Testi di riferimento:

Richard Sutton, Andrew Barto, Reinforcement Learning: An introduction. : MIT Press, 2018 Dimitri Bertsekas, Reinforcement Learning and Optimal Control. : Athena Scientific, 2019

Eventuali indicazioni sui materiali di studio:

I docenti metteranno a disposizione note ed appunti agli studenti.

SEMINARS AND OTHER ACTIVITIES

Titolare: Prof. WOLFGANG ERB

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, annuale

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: ; 3,00

STATISTICAL LEARNING (C.I.)

Titolare: Prof. ALBERTO ROVERATO

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Moduli del C.I.:

Statistical Learning 1 (Mod. A)

Statistical Learning 2 (Mod. B)

STATISTICAL LEARNING 1 (MOD. A)

Titolare: Prof. ALBERTO ROVERATO

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

Contenuti:

Part 1: - Data: summary statistics, displaying distributions; exploring relationships - Estimation: point estimation; the sampling distribution of an estimator; accuracy of estimation; interval estimation - Hypothesis testing - Likelihood: the likelihood, likelihood for several parameters - Estimation: maximum likelihood estimation; properties of maximum likelihood estimates

Attività di apprendimento previste e metodologie di insegnamento:

Lectures and Laboratories

Eventuali indicazioni sui materiali di studio:

Applications can be found in: - Nolan, D.A. & Speed, T. (2000). Stat Labs: Mathematical Statistics Through Applications. Springer. - Torgo, L. (2011). Data Mining with R: Learning with Case Studies. Chapman & Hall/CRC. Methods for specific fields of applications can be found in the following books: - Campbell, R.C. (1989). Statistics for Biologists (3rd ed.). Cambridge University Press. -Devore, J.L. (2000). Probability and Statistics for Engineering and the Sciences (5th ed.). Duxbury Press, Pacific Grove, CA. -Agresti, A. & Finlay. B. (2007). Statistical Methods for the Social Sciences (4th ed.). Prentice Hall

Testi di riferimento:

Lavine, M., Introduction to Statistical Thought. : None, 2013 Gareth, James, An Introduction to Statistical Learning. : Springer, 2013 Hastie, T., Tibshirani, R., and Friedman, J., The Elements of Statistical Learning: Data Mining, Inference, and Prediction. : Springer, 2001

STATISTICAL LEARNING 2 (MOD. B)

Titolare: Prof. ALBERTO ROVERATO

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

Contenuti:

Part 2: - Models : normal linear models; analysis of variance, analysis of covariance, inference for linear models; generalized linear models; inference for generalized linear models - Classification: logistic regression classifier, linear discriminant analysis, quadratic discriminant analysis - Cross validation and variable selection - Shrinkage methods: ridge regression, lasso regression - Networks: undirected graphical models

Attività di apprendimento previste e metodologie di insegnamento:

Lectures and Laboratories

Eventuali indicazioni sui materiali di studio:

Applications can be found in: - Nolan, D.A. & Speed, T. (2000). Stat Labs: Mathematical Statistics Through Applications. Springer. - Torgo, L. (2011). Data Mining with R: Learning with Case Studies. Chapman & Hall/CRC. Methods for specific fields of applications can be found in the following books: - Campbell, R.C. (1989). Statistics for Biologists (3rd ed.). Cambridge University Press. - Devore, J.L. (2000). Probability and Statistics for Engineering and the Sciences (5th ed.). Duxbury Press, Pacific Grove, CA. - Agresti, A. & Finlay, B. (2007). Statistical Methods for the Social Sciences (4th ed.). Prentice Hall

Testi di riferimento:

Lavine, M., Introduction to Statistical Thought. : None, 2013 Gareth James, An Introduction to Statistical Learning (2nd edition). : Springer, 2021 Hastie, Trevor J.; Tibshirani, Robert, The elements of statistical learning data mining, inference, and prediction. New York: Springer, 2009

STOCHASTIC METHODS

Titolare: Prof. MARCO FERRANTE

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 48A; 6,00

VISION AND COGNITIVE SYSTEMS

Titolare: Prof. LAMBERTO BALLAN

Mutuato da: Laurea magistrale in Computer Science (Ord. 2021)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligent systems

Tipologie didattiche: 32A+16L; 6,00

Prerequisiti:

Lo studente deve avere conoscenze di base di programmazione e algoritmi, così come di analisi matematica, probabilità e statistica, algebra lineare. È inoltre consigliabile una familiarità con concetti di base di apprendimento automatico.

Conoscenze e abilità da acquisire:

Questo corso insegna i concetti, i metodi e le tecnologie alla base della visione artificiale e dei sistemi cognitivi, incluso i moderni servizi cognitivi, vale a dire API e servizi tipicamente disponibili su cloud, che aiutano gli sviluppatori a creare applicazioni di intelligenza artificiale. Esempi di funzioni intelligenti che possono essere aggiunte ad un'applicazione tramite l'utilizzo di servizi cognitivi sono: il riconoscimento visuale; il rilevamento delle emozioni da video ed il riconoscimento facciale; comprensione linguistica e del parlato. Il corso insegna inoltre le competenze e le abilità specifiche necessarie per applicare tali concetti alla progettazione e all'implementazione di applicazioni di intelligenza artificiale. Gli studenti dovranno affrontare esercizi pratici in laboratorio informatico, in modo da provare l'applicazione delle conoscenze acquisite a piccoli esempi pratici.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso consiste in lezioni e esercizi in laboratorio informatico. Gli esercizi in laboratorio informatico consentono agli studenti di sperimentare, in diversi scenari operativi, le tecniche introdotte a lezione. In questo modo gli studenti possono verificare sperimentalmente i concetti appresi in classe, acquisire la capacità di applicare i concetti appresi e di esprimere un giudizio critico.

Contenuti:

Il corso comprende gli argomenti elencati di seguito: - Introduzione: Dalla cognizione umana all'intelligenza artificiale e ai sistemi cognitivi; breve introduzione ai paradigmi di intelligenza artificiale e apprendimento automatico; la rivoluzione dell'IA: attuali tendenze e applicazioni, le principali sfide. - Servizi cognitivi: Concetti basilari; servizi linguistici, vocali e di visione; principali provider e API (IBM Watson, AWS, Google Cloud); tecnologie abilitanti. - Apprendimento automatico ed applicazioni: Classificazione; introduzione al deep learning e all'apprendimento di rappresentazioni; fasi di addestramento e test; misure di valutazione; il bias negli algoritmi. - Visione ed elaborazione di immagini: Percezione nelle macchine; formazione dell'immagine, campionamento, filtraggio e operatori lineari; gradiente dell'immagine, edge e corner; progettare descrittori visuali efficaci (SIFT e feature basate sul gradiente); confronto tra immagini. - Riconoscimento visivo e oltre: "Insegnare ai computer a vedere": bag-of-feature, piramidi spaziali e pooling; apprendimento di rappresentazioni per la visione, reti neurali convoluzionali; R-CNN e segmentazione; descrizione di immagini, scenari multi-modalità e uno sguardo oltre al paradigma di apprendimento supervisionato. - Esercizi pratici: Cosa c'è nella scatola? Come costruire una pipeline di riconoscimento visivo; utilizzare i servizi cognitivi per il riconoscimento / comprensione delle immagini; combinare diversi servizi e modalità.

Modalità di esame:

Lo studente deve sviluppare, in accordo con il docente, un piccolo progetto applicativo. Inoltre, lo studente deve presentare una relazione scritta sul progetto svolto, in cui si discutono criticamente tutte le questioni trattate durante la sua realizzazione. L'esame consisterà prevalentemente in una breve presentazione e discussione del progetto svolto, in cui il docente potrà anche chiedere dettagli e/o altri contenuti visti a lezione.

Criteria di valutazione:

Il lavoro di progetto e l'esame orale saranno valutati sulla base dei seguenti criteri: a) conoscenza da parte dello studente dei concetti, dei metodi e delle tecnologie alla base dei servizi cognitivi (con particolare enfasi sulle tematiche di visione artificiale); b) capacità dello studente di padroneggiare la tecnologia di implementazione; c) capacità di sintesi, chiarezza e astrazione dello studente, come dimostrato dalla relazione scritta e dal progetto.

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

Le presentazioni mostrate durante le lezioni sono rese disponibili su Moodle come materiale di riferimento.

Curriculum: Data Science - Piano di studio Mathematics of Data Science**ADVANCED TOPICS IN DATA SCIENCE**

Titolare: Prof. LAMBERTO BALLAN

Periodo: I anno, annuale

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 24A; 3,00

DEEP LEARNING (MATICOLE DISPARI)

Titolare: Prof. ALESSANDRO SPERDUTI

Mutuato da: Laurea magistrale in Computer Science (Ord. 2021)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

DEEP LEARNING (MATICOLE PARI)

Titolare: Prof. ALESSANDRO SPERDUTI

Mutuato da: Laurea magistrale in Computer Science (Ord. 2021)

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

ENGLISH LANGUAGE B2 (PRODUCTIVE SKILLS)

Titolare: Prof. WOLFGANG ERB

Periodo: II anno, annuale

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: ; 3,00

FINANCIAL MATHEMATICS FOR DATA SCIENCE

Titolare: Prof. CLAUDIO FONTANA

Mutuato da: Laurea magistrale in Computational Finance (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Analisi stocastica (Propedeutico per gli studenti della laurea in matematica)

Conoscenze e abilità da acquisire:

The course presents some important models that are typically used in the banking industry. The students at the end should be familiar with pricing and hedging in both discrete and continuous time and they should be able to apply stochastic methods to the pricing of equity/forex/fixed income products

Attività di apprendimento previste e metodologie di insegnamento:

Lecture supported by tutorial, exercises and laboratory activities.

Contenuti:

Contenuti: The pricing problem in the binomial models Risk neutral pricing in the discrete time world European and American options in the binomial model. Arbitrage and risk neutral pricing in continuous time. Pricing of contingent claims in continuous time: the Black&Scholes formula. Black&Scholes via PDE and via Girsanov. Hedging and completeness in the Black&Scholes framework. Feynman-Kac formula and risk neutral pricing in continuous time. Put Call parity, dividends and static vs dynamic hedging. The Greeks and the Delta-Gamma hedging. Delta-Gamma-Vega neutral portfolios. Barrier options pricing in the Black&Scholes model. Quanto option pricing in the Black&Scholes model. Multi asset markets, pricing and hedging. Exchange options pricing in the multi-asset Black&Scholes model. Incomplete markets: quadratic hedging. Smile and skew stylized facts. Beyond the Black&Scholes model: stochastic volatility. The Heston model. Bonds and interest rates. Pre-crisis and multiple-curve frameworks. Short rate models, Vasicek, CIR, Hull-White models, affine models. Cap&Floor pricing in the short rate approaches.

Modalità di esame:

Final examination based on: Written and oral examination.

Criteri di valutazione:

Critical knowledge of the course topics. Ability to present the studied material.

Testi di riferimento:

Bjork Thomas, Arbitrage theory in continuous time. : Oxford Univ. Press, 2004 Lamberton D. and Lapeyre, B., Introduction to stochastic calculus applied to finance. : Cambridge Univ. Press, 2000

Eventuali indicazioni sui materiali di studio:

Lecture notes and reference books will be given by the teacher

FUNDAMENTALS OF INFORMATION SYSTEMS

Titolare: Prof. GIORGIO MARIA DI NUNZIO

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 96A; 12,00

GAME THEORY

Titolare: Dott. GIOVANNI PERIN

Mutuato da: Scuola Galileiana di Studi Superiori - Classe di Scienze Naturali

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Un corso basilare di teoria della probabilità.

Conoscenze e abilità da acquisire:

L'insegnamento prevede l'acquisizione delle seguenti conoscenze e abilità: - apprendere e padroneggiare concetti teorici di base e avanzati della teoria dei giochi e saper risolvere problemi generali multi-obiettivo multi-agente con tecniche della teoria dei giochi. - sapere applicare i concetti della teoria dei giochi a scenari pratici in ambito informatico e di analisi dei dati, soprattutto per metodologie con intelligenza computazionale, multi-agente e distribuite.

Attività di apprendimento previste e metodologie di insegnamento:

Lezioni convenzionali con il supporto di slide. Prevista interazione su piattaforma moodle. Possibile utilizzo di video registrati.

Contenuti:

Concetti base di teoria dei giochi Utilità, mercato, fattore di sconto Giochi statici in forma normale Dominanza, Equilibri di Nash Efficienza, prezzo dell'anarchia Giochi a somma zero, giochi minimax Strategie miste, equilibri misti Teorema di Nash, il teorema minimax The tragedy of the commons Giochi dinamici Strategie e sottogiochi Backward utility Equilibri di Stackelberg Giochi ripetuti, collaborazione Duopoli dinamici, collusione Cooperazione, pricing Informazione incompleta/imperfetta Giochi bayesiani, signaling, beliefs Principio di rivelazione Teoria dei giochi assiomatica Fictitious play Best response dynamics Ottimizzazione distribuita Game theory algoritmica Calcolo, complessità, e completezza dell'equilibrio Aste, bargaining Aste di primo e secondo prezzo, criterio VCG (cenni) Giochi cooperativi, il nucleo, il valore di Shapley (cenni) Allocazione delle risorse Utilità, scelte e paradossi Giochi potenziali, coordinazione Algoritmi bio-inspired Giochi evolutivi Reti cognitive Selfish routing Age of information Sistemi multi-input con teoria dei giochi

Modalità di esame:

Test scritto obbligatorio a libro aperto, con diversi problemi di game theory su argomenti toccati durante il corso. Questo determina il punteggio base della votazione per un massimo di 27 punti. I frequentanti possono in aggiunta sviluppare un progetto in gruppi di 1-3 persone, su argomenti del corso. Adesione al progetto e argomento dello stesso sono concordati con il docente durante il corso. Il progetto va consegnato entro un mese dalla fine delle lezioni. Se lo scritto è sufficiente (maggiore di 15) si può registrare il voto finale: - senza progetto: punteggio conseguito allo scritto + 3 punti - con progetto: punteggio

conseguito allo scritto + valutazione del progetto che va da 0 a 8 punti

Criteri di valutazione:

Ogni domanda nei test scritti ha valutazione numerica comunicata in anticipo. Il punteggio finale dello scritto è la somma numerica dei punteggi individuali delle domande (max 27). Studenti che non svolgono il progetto ricevono ulteriori 3 punti per arrivare fino a un massimo di 30. Il progetto viene invece valutato da 0 a 8 punti, sommati al punteggio dello scritto. Un punteggio di 30 e lode è assegnato agli studenti il cui punteggio totale è maggiore o uguale di 31. Nella valutazione di ogni domanda scritta vengono tenuti in considerazione: - la pertinenza, la correttezza, e la completezza della risposta; - l'utilizzo appropriato delle terminologie, metodologie, e rappresentazioni formali tipiche della teoria dei giochi - l'acquisita capacità di problem solving - la capacità di discussione e verifica ex-post della soluzione trovata Nella valutazione del progetto (se presente) vengono tenuti in considerazione: - l'originalità della proposta e la pertinenza con le tematiche del corso - la capacità di lavoro di gruppo e la presenza di singoli contributi attribuibili ai partecipanti al progetto - la capacità di trarre conclusioni significative dal punto di vista scientifico grazie alle metodologie apprese nel corso

Testi di riferimento:

S. Tadelis, Game Theory: An Introduction. : Princeton, 2013 A. MacKenzie, L. Da Silva, Game Theory for Wireless Engineers. Morgan & Claypool: 2006, N. Nisan, T. Roughgarden, E. Tardos, V. V. Vazirani, Algorithmic game theory. : Cambridge Univ. Press, 2007 R. Lucchetti, A Primer in Game Theory. Bologna: Esculapio, L. Badia, T. Marchioro, Game theory. A handbook of problems and exercises. Bologna: Esculapio, 2022

Eventuali indicazioni sui materiali di studio:

Il corso segue la trattazione standard in ambito accademico per la presentazione delle basi della teoria dei giochi. Pertanto, il libro di Tadelis (integrato con materiale fornito a lezione/moodle) è sufficiente come riferimento generale, ma è equivalente ad ogni altro libro di introduzione alla teoria dei giochi. Per gli esercizi, si consiglia il Badia-Marchioro che contiene esercizi analoghi a quelli d'esame. Per le applicazioni e progetti, si può fare riferimento al libro di MacKenzie-DaSilva, ma va bene ogni materiale di ricerca trovato in rete.

HIGH DIMENSIONAL PROBABILITY FOR DATA SCIENCE

Titolare: Prof. ALBERTO CHIARINI

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

INTERNSHIP

Titolare: Prof. TOMASO ERSEGHE

Periodo: Il anno, annuale

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: ; 18,00

LAW AND DATA

Titolare: Dott.ssa FIORELLA DAL MONTE

Periodo: Il anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

Prerequisiti:

No prerequisites

Conoscenze e abilità da acquisire:

The course aims to introduce non-law students to a proper understanding of the main legal issues related to the processing of data, personal and non. The first part of the course aims to enable students to approach EU personal data protection regulation. In the second part, instead, students will reflect on the main problems related to the use of data-intensive technologies (big data and artificial intelligence) and the technical and legal solutions now debated.

Attività di apprendimento previste e metodologie di insegnamento:

Classes Seminars Workshops Preassigned readings.

Contenuti:

All the info about the course are on Moodle - Introduction to Law and Legal Studies - Introduction to the EU Law - Introduction to the EU GDPR - The concept of data; personal, sensitive and economic data; big data - Property of data, choices in the management of data - The right to be forgotten - Civil and criminal aspects of profiling activity - Automatic data processing, human responsibilities - The Data Protection Officer and DP Authorities - Civil and criminal protection of privacy - Sanctioning powers and system - Open Data for the public interest - Big data (collection, analysis, processing) and their influence on fundamental rights - Digital Surveillance - Facial Recognition: Open Issues - Disinformation - Artificial Intelligence in the EU law

Modalità di esame:

Written Exam

Criteri di valutazione:

The grading scale used to assess the students is the Italian one, with the highest score of 30/30 and a minimum score of 18/30 (sufficient) (info: here). The students will be graded according to their level of theoretical and practical knowledge of the fields covered throughout the course and their capacity to critically reflect on the most contentious legal issues on data-intensive technologies.

Testi di riferimento:

Eventuali indicazioni sui materiali di studio:

The course has no official textbooks. Students can study on their notes and the additional material provided by the instructor. Nevertheless, here are some helpful handbooks to approach some modules. These books are not mandatory, and students have sole discretion to refer them. ? Mireille Hildebrandt (2020). Law for Computer Scientists and Other Folk, OUP (open access: [here](#)) – especially chapters 2-3-4-5-9-10 ? Paul Voigt, Axel von dem Bussche (2017). The EU General Data Protection Regulation (GDPR). A practical guide, Springer (unipd access: [here](#)) ? European Fundamental Rights Agency (2018). Handbook on European data protection law, Luxembourg (open access: [here](#)) ? Karen Yeung, Martin Lodge (2019). Algorithmic Regulation, OUP (Public Law Dept. Library) – especially chapters 2-3-4-6-7-11

MACHINE LEARNING

Titolare: Dott. DAVIDE RIGONI

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Sono richieste conoscenze di base di Analisi Matematica, Probabilità e Statistica, Algebra Lineare, e di Programmazione.

Conoscenze e abilità da acquisire:

Il corso presenta i concetti e gli algoritmi fondamentali dell'Apprendimento Automatico, cioè quella classe di tecniche ed algoritmi che a partire dai dati consentono di acquisire nuova conoscenza. Questi sono particolarmente utili per tutti quei problemi per cui è impossibile, o molto difficile, pervenire ad una formalizzazione utilizzabile per la definizione di una soluzione algoritmica ad hoc. Esempi di tali problemi sono i compiti percettivi, come il riconoscimento visivo o il riconoscimento del parlato, e problemi in cui i dati sono affetti da rumore o sono incompleti. Sono previste varie esercitazioni in laboratorio informatico, al fine di consentire allo studente di sperimentare le conoscenze acquisite mediante la loro applicazione a piccoli esempi pratici. Il corso insegna inoltre le competenze necessarie per applicare e valutare modelli di apprendimento automatico in contesti applicativi reali.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso consiste in lezioni e esercizi in laboratorio informatico. Gli esercizi in laboratorio informatico consentono agli studenti di sperimentare, in diversi scenari operativi, le tecniche introdotte a lezione. Gli studenti possono così verificare sperimentalmente i concetti appresi in classe, ed acquisire la capacità di applicarli criticamente di fronte ad un problema concreto.

Contenuti:

Il corso tratta gli argomenti elencati di seguito: - Introduzione Intelligenza Artificiale ed Apprendimento Automatico, quando e perché utilizzare tecniche di Apprendimento Automatico; I diversi paradigmi di Apprendimento Automatico, caratteristiche ed applicazioni; Gli ingredienti Fondamentali dell'Apprendimento Automatico. - Apprendimento Supervisionato: le fondamenta ed i modelli base Modelli di Regressione Lineare; spazio delle ipotesi, rappresentazione e definizione della funzione di costo; ottimizzazione e discesa del gradiente. Modelli di Classificazione Lineare; Regressione Logistica; Regolarizzazione e model selection. - Bontà di un modello e sua valutazione Bias-Variance Tradeoff e come affrontare problemi di overfitting ed underfitting; minimizzazione del rischio e teoria dell'apprendimento; misure di valutazione delle performance, esempi ed applicazioni; come effettuare debugging e diagnosticare il comportamento di un sistema di Apprendimento Automatico. - Apprendimento Supervisionato: reti neurali e modelli avanzati Reti Neurali Artificiali; perceptron, reti multistrato e cenni di deep learning; apprendimento dei parametri in reti neurali, backpropagation e discesa di gradiente. Support Vector Machines; metodi kernel e classificazione non lineare. Paradigmi "alternativi": metodi non-parametrici (k-NN) e loro applicazioni; alberi di decisione e random forest. - Apprendimento non Supervisionato Clustering: KMeans e cenni agli approcci principali; PCA e riduzione di dimensionalità. - Introduzione ai sistemi di Raccomandazione e alla loro applicazione

Modalità di esame:

L'esame verrà svolto con una prova scritta.

Criteri di valutazione:

La valutazione dello studente si basa sull'accertamento dell'apprendimento degli argomenti trattati nel corso e sulle sue capacità di analisi.

Testi di riferimento:

Bishop, Christopher M., Pattern recognition and machine learning. New York: Springer, 2006

Eventuali indicazioni sui materiali di studio:

Le presentazioni mostrate durante le lezioni e link di approfondimento sono resi disponibili su Moodle come materiale di riferimento.

MACHINE LEARNING FOR HUMAN DATA

Titolare: Prof. MICHELE ROSSI

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Il corso richiede conoscenze in Teoria della Probabilità (variabili aleatorie, probabilità condizionata e formula di Bayes, distribuzioni di probabilità discrete e continue) e un minimo di conoscenza di linguaggi di programmazione (MatLab, Python). Anche se non strettamente necessarie, conoscenze nell'ambito di algebra lineare (es, spazi vettoriali e decomposizione ai valori singolari) e dell'analisi numerica dei segnali (trasformate di Fourier discrete) sono utili. Il Professore effettuerà dei ripassi per le tecniche e teorie di cui sopra, qualora necessario. Materiale e referenze per il ripasso personale da parte dello studente di tali tecniche verranno altresì forniti. Inoltre, anche se conoscenze pregresse sono sicuramente utili per lo studente, il corso è concepito in modo da non dipendere criticamente da esse. Benchè non obbligatorio, un esame di base di "Machine Learning" è utile per una migliore comprensione del corso.

Conoscenze e abilità da acquisire:

1. Conoscere i principali algoritmi per la classificazione di dati multidimensionali, i loro pro e contro, le metriche per la loro valutazione 2. Conoscere le

principali tecniche per l'apprendimento non-supervisionato (vector quantization), le loro prestazioni, i vantaggi e il loro utilizzo in seno a problemi reali nell'ambito dei biosegnali 3. Conoscere le principali tecniche di modellizzazione di serie temporali multivariate e il loro utilizzo per la soluzione di problemi reali 4. Conoscere i principi e le tecniche dell'apprendimento supervisionato con particolare riferimento alle reti neurali (feed forward, convoluzionali e ricorrenti), la loro programmazione in Python, e il loro utilizzo in seno a problemi reali 5. Conoscere i principali ambiti di applicazione, delle tecniche ai punti 1, 2, 3 e 4 e come queste tecniche sono state utilizzate per risolvere problemi nell'ambito "human data" 6. Acquisire la sensibilità necessaria per saper discernere e utilizzare i modelli ai punti 1, 2, 3, e 4 7. Essere in grado di risolvere un problema di analisi dei dati e: 1) sintetizzarne la soluzione in un documento (project report) scritto in modo professionale, 2) presentare il lavoro svolto oralmente nella forma di "presentazione stile conferenza" con dimostrazione del software scritto allo scopo 8. Essere in grado di utilizzare e implementare al computer gli algoritmi ai punti 1, 2, 3 e 4 tramite il linguaggio Python

Attività di apprendimento previste e metodologie di insegnamento:

Il corso è articolato come segue: - Didattica frontale (36 ore): il docente presenterà gli argomenti presenti nel syllabus delle lezioni utilizzando principalmente delle slide. Particolari passaggi matematici verranno presentati tramite l'ausilio della lavagna. --- Una lezione verrà dedicata alla presentazione del progetto del corso, del set di dati da utilizzare e delle modalità di esame --- Il corso è progettato, per ogni sezione, in modo da presentare prima una trattazione teorica dei metodi e dei modelli, per poi descrivere alcune applicazioni di riferimento, spiegando nel dettaglio come i metodi vengono usati / adattati nei vari ambiti, commentando scelte architetturali e le relative prestazioni - Attività di laboratorio (12 ore): sei lezioni verranno dedicate all'attività di laboratorio, nella quale le tecniche sviluppate durante le lezioni teoriche verranno implementate e caratterizzate sperimentalmente utilizzando il linguaggio Python. Tutto il materiale didattico presentato / utilizzato (slides e software) sarà reso disponibile sul sito Moodle del corso.

Contenuti:

Parte I – Introduzione (2 ore) - Introduzione al corso, modalità di esame, orari del docente, etc. - Applicazioni: salute, servizi che includono il tracciamento di attività, applicazioni di sicurezza e emergenza, analisi di movimento, natural language processing Part II – Tecniche di quantizzazione vettoriale (18 ore) - Quantizzazione Vettoriale (QV): -- Significato, utilità, metriche -- Clustering 1: K-means, soft K-means, Expectation Maximization (EM) -- Clustering 2: X-means, DBSCAN (density based clustering) - Algoritmi per la QV non-supervisionata: --- Mappe Auto-Organizzanti (SOM), Reti Neurali "Gas" (GNG): teoria e algoritmi - Applicazioni a segnali biometrici quasi-periodici (ECG): -- Pre-processamento del segnale, normalizzazione, segmentazione --- Apprendimento di dizionari: concetti, architetture --- Rappresentazione efficiente di segnali ECG: descrizione di tecniche dello stato dell'arte --- Esempi di architetture per l'apprendimento non-supervisionato GNG-based per segnali ECG --- Architettura finale del sistema e risultati numerici Part III - Reti Neurali "Deep" (16 ore) - Review: tecniche di discesa del gradiente e concetti generali (apprendimento supervisionato, modelli di costo, etc.) - Review: Reti neurali "feed forward": modelli, allenamento delle reti, algoritmo di "back-propagation" - review - Reti neurali convoluzionali (CNN): struttura, blocchi costituenti, allenamento, batch normalization - Architetture autoencoder per l'estrazione automatica di feature -- denoising autoencoder - Reti ricorrenti (RNN) per l'analisi di sequenze temporali. Esempio: natural language processing -- Autoencoder per sequenze temporali basati su RNN - Residual convolutional neural networks (ResNets) - Meccanismi di attenzione per reti CNN (immagini) e RNN (sequenze temporali) - Architetture Transformer: spiegazione passo passo dei principi di funzionamento, dei blocchi architetturali e dei meccanismi per il loro training. Esempi applicativi e trend Part IV: Lezioni di laboratorio (12 ore) Le lezioni di laboratorio saranno effettuate in modo guidato, presentando allo studente codice pre-scritto e funzionante, ma allo stesso tempo soffermandosi su ogni blocco del codice, caratterizzandone il funzionamento e i dati che esso ritorna in uscita. Le lezioni si incentreranno sulla comprensione del linguaggio Python e dei vari tools (Keras e TensorFlow) per la definizione e l'allenamento di reti neurali. Le reti neurali "feed forward", quelle convoluzionali e ricorrenti, verranno pienamente caratterizzate descrivendo i blocchi che le compongono, la loro programmazione e la creazione delle varie architetture neurali. Queste verranno infine allenate tramite diversi algoritmi che sfruttano la discesa del gradiente. Le reti trattate verranno poi testate con opportuni dataset. In estrema sintesi, gli argomenti trattati sono: - Introduzione alla programmazione con il linguaggio Python - Reti neurali con struttura "feed forward" - Reti neurali convoluzionali (CNN) - Reti neurali ricorrenti (RNN) - Meccanismi avanzati (implementazione): batch normalization, residual networks (ResNets), autoencoders, inception layers

Modalità di esame:

Questo è un corso di "machine learning" avanzato e applicato a problematiche reali. Pertanto, la verifica finale dell'apprendimento dello studente verterà su un progetto, che coinvolgerà le seguenti fasi di lavoro: 1. Il Professore assegnerà agli studenti un problema da risolvere, individuando un dataset aperto, pubblicamente fruibile e utilizzabile allo scopo. Il problema verrà descritto dal docente tramite un'apposita lezione, nella quale si spiegheranno altresì le modalità di presentazione dei risultati finali, via 1) una relazione finale, in forma scritta, 2) una presentazione, in forma orale 2. Gli studenti si distribuiranno in gruppi, con un massimo di due persone per gruppo di lavoro, e inizieranno a lavorare al progetto assegnato. La scelta della tecnica da utilizzare, il pre-trattamento dei dati per ottenere delle feature informative, etc. saranno tutti dettagli da risolvere, a discrezione dello studente. Il docente si renderà disponibile per seguire i gruppi nelle varie fasi di lavoro 3. Ogni gruppo risolverà il problema proposto secondo la tecnica prescelta e presenterà: 1) una relazione finale, 2) una presentazione al docente in forma orale dove si descriverà, il problema affrontato, i modelli e tecniche utilizzati, il codice realizzato allo scopo, i risultati ottenuti. È inoltre apprezzata la dimostrazione del funzionamento del codice realizzato da parte degli studenti Il voto finale verrà proposto dal docente dopo un'attenta valutazione della relazione scritta al punto 1) e della presentazione in forma orale al punto 2).

Criteri di valutazione:

I criteri di valutazione con cui verrà effettuata la verifica delle conoscenze e delle abilità attese, saranno: 1. Completezza delle conoscenze acquisite 2. Capacità di analisi di un problema reale attraverso le tecniche presentate nel corso 3. Proprietà nella terminologia tecnica usata, sia scritta che orale 4. Originalità e indipendenza nella identificazione della soluzione scelta per la soluzione del progetto 5. Competenza e coerenza nell'interpretazione del significato dei risultati ottenuti / qualità degli stessi 6. Abilità nell'utilizzo degli strumenti informatici nello studio del problema assegnato 7. Qualità dell'esposizione orale 8. Qualità dell'esposizione scritta

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

Report tecnici, articoli scientifici, software (Python code) e tutto il materiale necessario per un corretto apprendimento dei concetti esposti saranno resi disponibili dal docente tramite la pagina Moodle del corso. - Le slide e le registrazioni delle lezioni verranno rese disponibili tramite la pagina Moodle del corso - Altri libri utili (non obbligatori) sono: Per un ripasso di concetti di algebra lineare: - J. R. Magnus and H. Neudecker, "Matrix Differential Calculus with Applications in Statistics and Econometrics," Wiley, 1999. Per un ripasso dei concetti matematici propedeutici al machine learning: - Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong, "Mathematics for Machine Learning," Cambridge University Press, 2020.

MATHEMATICAL CELL BIOLOGY

Titolare: Prof. MORTEN GRAM PEDERSEN

Mutuato da: Scuola Galileiana di Studi Superiori - Classe di Scienze Naturali

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

MATHEMATICAL MODELS AND NUMERICAL METHODS FOR BIG DATA

Titolare: Prof. WOLFGANG ERB

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Conoscenze di base di analisi matematica e di algebra lineare. Conoscenze di base di algebra lineare e delle matrici. - Spazi vettoriali, indipendenza lineare, base in spazi vettoriali - Prodotto scalare e ortogonalita' - Tipi di matrici: diagonale, triangolare, simmetrica, definita positiva - Spettro di una matrice: autovalori e autovettori - Fattorizzazione di una matrice: LU, Cholesky, QR - Risoluzione di sistemi lineari Conoscenze di base di analisi matematica. - Norme e metriche in spazi vettoriali - Continuita' di funzioni e di operatori - Successioni e convergenza Conoscenze di base di programmazione (Matlab o Python).

Conoscenze e abilità da acquisire:

In questo corso si apprenderanno gli strumenti matematici e computazionali d'avanguardia per l'analisi numerica di megadati soprattutto in vista delle applicazioni nel machine learning. Usando pacchetti di software moderni per dati grandi e sparsi (principalmente Matlab o Python), si impara come implementare questi metodi in maniera efficiente e come applicarli in problemi di vita reale.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso si baserà principalmente su lezioni frontali in aula ed esercitazioni. Dispense e slides verranno rese disponibili sulla pagina web del corso.

Contenuti:

1. Ranking with eigenvectors • Recapitulation of important results from linear algebra • The vector iteration: calculation of dominant eigenpairs • Ranking of web pages: PageRank and Hits. 2. Numerical methods for large scale linear systems • Krylov subspace methods for large eigenvalue problem: Arnoldi and Lanczos • Krylov subspace methods for sparse systems of equations: GMRES and MINRES • Krylov subspace methods for the computation of matrix functions 3. Spectral graph theory • Graphs, the graph Laplacian, graph signals, the Cheeger constant • The graph Fourier transform, graph convolution, filtering and decomposition of graph signals • Dimensionality reduction with Laplacian eigenmaps • Centralities on graphs 4. Clustering algorithms • k-center clustering and hierarchical k-center clustering • The k-means algorithm • Spectral clustering 5. Low rank matrix approximations • Singular value decomposition (SVD): basic properties, numerical computation and best rank-k approximation • Principal component analysis and dimensionality reduction • Latent factor models and recommender systems • Matrix factorizations beyond SVD: Funk SVD and nonnegative matrix factorizations. 6. Multiway Data Analysis • Tensors and tensor decomposition • Higher order singular value decomposition (HOSVD): definition and numerical computation • Applications: Face Recognition Using tensor SVD and Tensor Data Fusion • Kernel methods for data analysis

Modalità di esame:

Esame scritto alla fine del corso. Esercizi a casa durante il periodo delle lezioni forniranno un ulteriore feedback e crediti aggiuntivi per l'esame finale.

Criteri di valutazione:

La prova scritta alla fine del corso mira a verificare la comprensione dei fondamenti teorici e delle competenze pratiche apprese durante il corso. La valutazione nella prova finale si baserà: - sulla comprensione dei modelli matematici e la teoria vista a lezione - sulla capacità di utilizzare i metodi numerici presentati in aula in maniera autonoma per la risoluzione di problemi concreti.

Testi di riferimento:

Eldén, Lars, Matrix methods in data mining and pattern recognition. : Society for Industrial and Applied Mathematics, 2019 Saad, Yousef, Numerical methods for large eigenvalue problems. Philadelphia: SIAM, 2011 Daniela Calvetti; Erkki Somersalo, Mathematics of data science : a computational approach to clustering and classification : pbk. : Society for Industrial and Applied Mathematics, 2021

Eventuali indicazioni sui materiali di studio:

Dispense estese, slides e codici verranno fornite sul sito web del corso

NETWORK SCIENCE

Titolare: Prof. TOMASO ERSEGHE

Mutuato da: Scuola Galileiana di Studi Superiori - Classe di Scienze Naturali

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

OPTIMIZATION FOR DATA SCIENCE

Titolare: Prof. FRANCESCO RINALDI

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

PROCESS MINING

Titolare: Prof. MASSIMILIANO DE LEONI

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Conoscenze di base di algoritmi, strutture dati e programmazione, come acquisiti nel corso "Fundamental of Information Systems" o nei corsi di laurea in Informatica o affini.

Conoscenze e abilità da acquisire:

Oggigiorno, tutte le medie e grandi imprese seguono procedure e processi standardizzati per fornire beni e servizi, al fine di garantire livelli di servizio e di efficienza più alti. La modellazione formale di questi processi viene utilizzata come input per sistemi informativi (per es. sistemi SAP o ERP) che automaticamente si autoconfigurano per funzionare in accordo con tali modelli di processo. Lo scopo del corso è quello di introdurre e studiare concetti, metodi e tecniche che riguardano la modellazione, l'analisi e il miglioramento di processi di business. Processi più efficienti portano a sistemi più efficienti, che a loro volta permettono alle organizzazioni di fornire migliori servizi e beni di più alta qualità. Il corso può concettualmente essere suddiviso in tre parti: 1) In una prima parte, il corso introduce i principi e i linguaggi alla base della modellazione dei processi. 2) In una seconda parte, la più estesa, il corso discute le tecniche di analisi dei processi aziendali basate sul Process Mining. Process Mining si pone l'obiettivo di utilizzare lo storico dei dati transazionali delle aziende (i log degli eventi) per analizzare e migliorare i processi aziendali. Il Process Mining utilizza le informazioni nei log degli eventi per scoprire come i processi vengono REALMENTE eseguiti, svelando le violazioni di protocolli e regolamenti, le inefficienze e tutte le criticità che mettono a rischio le esecuzioni stesse dei processi. Mentre le tecniche tradizionali di Business Intelligence si basano su interviste e questionari con gli attori dei processi, Process Mining è parte integrante di Data Science: la soggettività degli attori viene sostituita dall'oggettività delle conclusioni estratte dai dati effettivi di processo. 3) L'obiettivo ultimo del Process Mining è predisporre azioni concrete che possano portare a migliorare i processi, sia dal punto di vista di aderenza alle normative vigenti, sia dal punto di vista dell'efficienza. Nella terza parte del corso, gli "insight" ottenuti tramite Process Mining vengono utilizzati per costruire modelli di simulazioni dei processi aziendali che siano "gemelli digitali" dei processi reali. Tali "gemelli digitali" rispecchiano il comportamento dei processi reali, in tutti gli aspetti (caratteristiche e vincoli temporali, decisioni, dati manipolati, risorse, ecc.). Il vantaggio di avere modelli di simulazione (i gemelli digitali) risiede nel fatto che sia è possibile simulare variazioni di tali processi e misurare gli impatti sulle performance dei processi, senza dover veramente cambiare il processo reale, riducendo i costi e rischi di eseguire processi alternativi ancora più costosi e meno efficienti degli originali.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso consiste di lezioni, esercitazioni, sessioni di laboratorio e seminari. Le sessioni di laboratorio daranno l'opportunità di "sporcarsi le mani" e verranno prevalentemente utilizzati dataset reali per meglio comprendere la complessità di eseguire un progetto di Process Mining e di lavorare con dati di processo reali. I seminari verranno organizzati da aziende nazionali e internazionali del settore di Process Mining

Contenuti:

Il corso affronterà i seguenti argomenti: 1. MODELLAZIONE DEI PROCESSI TRAMITE RETI DI PETRI - Concetti di base di reti di Petri - Uso delle reti di petri per la modellazione dei processi aziendali - Analisi strutturale di reti di Petri - Correttezza di modelli di processi aziendali basati su reti di Petri: Principi e Verifica. - Uso di Woped come software per modellare e verificare la correttezza dei modelli di processo 2. PROCESS MINING - Introduzione su Process Mining e il log degli eventi - Tecniche di Base per Process Discovery e loro Limiti - Tecniche Avanzate per Process Discovery: Heuristic Miner and Region Miner - Verifica di Conformità basata su Token Replay - Verifica di Conformità basata sugli Allineamenti - Il Mining delle Prospettive Aggiuntive su Decisioni, Tempistiche e Risorse - Analisi delle Reti Social - Uso di software open-source and commerciali per l'analisi dei processi aziendali tramite Process Mining 3. SIMULAZIONE DEI PROCESSI AZIENDALI - Principi e Metodologie per la simulazione: Definizione di intervalli di Warm-up e Cool-down, gestione della stocasticità della simulazione tramite simulazioni multiple ed indipendenti, analisi dei risultati. - Definizione degli elementi di uno scenario di simulazione di processi aziendali: arrival rate, branching probabilities, resource pools, distribuzioni di probabilità delle durate delle attività, - Uso di BPMN per la specifica di modelli di simulazione di processi aziendali - Uso di Process Mining per la generazione di modelli di simulazione - Metodologie per la definizione scenari "What-if" - Uso di BIMP come software per la simulazione di processi aziendali.

Modalità di esame:

Esame scritto e Progetto

Criteri di valutazione:

Il lavoro di progetto e l'esame scritto saranno valutati sulla base dei seguenti criteri: i) conoscenza dello/a studente/ssa di concetti, metodi e tecnologie; ii) abilità dello/a studente/ssa di padroneggiare le tecnologie di implementazione; iii) le capacità di sintesi, precisione, chiarezza e astrazione dello/a studente/ssa dimostrate nell'esame scritto e nella presentazione di progetto. Il voto finale è ottenuto come la somma pesata dei voti dell'esame scritto (60%) e del progetto (40%).

Testi di riferimento:

Marlon Dumas, Marcello La Rosa, Jan Mendling, Hajo A. Reijers, Fundamentals of business process management. Berlin: Springer, 2018 Aalst, Wil M. P. : van der, Modeling business processes a Petri net-oriented approach. Cambridge: MIT Press, 2011 van der Aalst, Wil, Process Mining Data Science in Action. Berlin, Heidelberg: Springer Berlin Heidelberg Imprint: Springer, 2016

Eventuali indicazioni sui materiali di studio:

Tre testi verranno utilizzati per le tre parti, come indicato nel box "Testi di riferimento" sottostante Questi sono complementati da (1) un eserciziaro e da (2) documenti relativi ad esercitazioni autoguidate per la parte di laboratorio. Sia (1) che (2) saranno messi a disposizione su Moodle.

SEMINARS AND OTHER ACTIVITIES

Titolare: Prof. WOLFGANG ERB

Periodo: Il anno, annuale

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: ; 3,00

STATISTICAL LEARNING (C.I.)

Titolare: Prof. ALBERTO ROVERATO

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Moduli del C.I.:

Statistical Learning 1 (Mod. A)

Statistical Learning 2 (Mod. B)

STATISTICAL LEARNING 1 (MOD. A)

Titolare: Prof. ALBERTO ROVERATO

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

STATISTICAL LEARNING 2 (MOD. B)

Titolare: Prof. ALBERTO ROVERATO

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

STATISTICAL METHODS FOR HIGH DIMENSIONAL DATA

Titolare: Prof. BRUNO SCARPA

Periodo: II anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Statistical learning, Stochastic methods

Conoscenze e abilità da acquisire:

This course aims at introducing the students to the main statistical features and concepts underlying the analysis of high dimensional data, as well as providing statistical solutions to problems arising when analysing real data on many different fields (business, society, medicine, psychology, physics, etc).

Attività di apprendimento previste e metodologie di insegnamento:

Class lessons. Laboratory sessions using R

Contenuti:

Every year some of the following topics will be presented, according also to the preferences of the students. 1. REGRESSION MODELS FOR HIGH-DIMENSIONAL DATA 1.1 Incremental algorithms with limited memory, stochastic gradient descent, inference 1.2 Sparsity, penalization inducing sparsity 1.3 Recall of Lasso and Elastic-Net for GLM 1.4 Extensions: adaptation, fusion, dealing with categorical variables 1.5 Group LASSO 1.6 Non-convex penalties 2. STATISTICAL ANALYSIS OF NETWORK DATA 2.1 Introduction to network structures of data 2.2 Network and nodes indicators 2.3 Community detection 2.4 Basics statistical models and inference (Erdos-Renyi, p_1 , ERGM) 2.5 Bayesian models (Stochastic block models, Latent space models) 3. COMPUTATIONAL ISSUES

Modalità di esame:

Practical and oral exams

Criteri di valutazione:

Students will be evaluated according to their level of knowledge of the key concepts in analysing high dimensional data and their ability to apply them to real cases.

Testi di riferimento:

Hastie, Trevor; Tibshirani, Robert; Wainwright, Martin, Statistical Learning with Sparsity. Philadelphia, PA: Chapman and Hall/CRC, 2015

Eventuali indicazioni sui materiali di studio:

The teacher in charge will provide lecture notes, exercises and scientific papers

STOCHASTIC METHODS

Titolare: Prof. MARCO FERRANTE

Periodo: I anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Nozioni di base di calcolo differenziale e integrale, algebra lineare e calcolo delle probabilità.

Conoscenze e abilità da acquisire:

Lo scopo del corso è di introdurre metodi e concetti di Calcolo delle Probabilità e di alcune classi di Processi Stocastici che hanno un forte impatto come strumenti algoritmici, computazionali e nello studio delle reti.

Attività di apprendimento previste e metodologie di insegnamento:

Lezioni frontali integrate da materiale online. Esercitazioni in aula

Contenuti:

1. Richiami di Calcolo delle Probabilità. • spazi di probabilità, sigma-algebre, proprietà della probabilità • probabilità condizionata e indipendenza • variabili aleatorie, valor atteso e valor atteso condizionale • Legge dei grandi numeri e Chernoff Bounds • approssimazione di distribuzioni di probabilità. 2. Catene di Markov e passeggiate aleatorie • Catene di Markov a tempo discreto: definizione e classificazione degli stati • distribuzioni stazionarie e risultati di convergenza. 3. Martingale a tempo discreto • Proprietà del valore atteso condizionato • Definizione di martingala, sopra- e sotto martingala. • Tempi di arresto

Modalità di esame:

Esame scritto

Criteri di valutazione:

Il voto finale è basato sul risultato della prova scritta, il cui scopo principale è verificare la capacità di usare in modo corretto ed efficiente le tecniche esposte, applicandole a problemi concreti.

Testi di riferimento:

Paolo Dai Pra, Stochastic Methods for Data Science. : ,

Eventuali indicazioni sui materiali di studio:

Il docente fornirà delle dispense ed esercizi

Curriculum: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

ADVANCED MACHINE LEARNING

Titolare: da definire

Periodo: Il anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

Tipologie didattiche: 40A+16L; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

ADVANCED MULTIVARIATE ANALYSIS

Titolare: da definire

Periodo: Il anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

Tipologie didattiche: 42A+14L; 6,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

ADVANCED STATISTICAL MODELING

Titolare: da definire

Periodo: Il anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

Tipologie didattiche: 45A; 6,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

BIG DATA COMPUTING

Titolare: Prof. FRANCESCO SILVESTRI

Mutuato da: Scuola Galileiana di Studi Superiori - Classe di Scienze Naturali

Periodo: I anno, 2 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

Tipologie didattiche: 48A; 6,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

BIOINFORMATICS AND STATISTICAL GENETICS**Titolare:** da definire**Periodo:** Il anno, 1 semestre**Indirizzo formativo:** TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)**Tipologie didattiche:** 18A+30L; 6,00**Prerequisiti:**

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

COMPLEX AND SOCIAL NETWORKS**Titolare:** da definire**Periodo:** Il anno, 1 semestre**Indirizzo formativo:** TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)**Tipologie didattiche:** 54A; 6,00**Prerequisiti:**

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

FUNDAMENTALS OF INFORMATION SYSTEMS

Titolare: Prof. GIORGIO MARIA DI NUNZIO

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

Tipologie didattiche: 96A; 12,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

INFORMATION RETRIEVAL AND RECOMMENDER SYSTEMS

Titolare: da definire

Periodo: Il anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

Tipologie didattiche: 26A+26L; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

LAW AND DATA

Titolare: Dott.ssa FIORELLA DAL MONTE

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

Tipologie didattiche: 48A; 6,00

Prerequisiti:

No prerequisites

Conoscenze e abilità da acquisire:

The course aims to introduce non-law students to a proper understanding of the main legal issues related to the processing of data, personal and non. The first part of the course aims to enable students to approach EU personal data protection regulation. In the second part, instead, students will reflect on the main problems related to the use of data-intensive technologies (big data and artificial intelligence) and the technical and legal solutions now debated.

Attività di apprendimento previste e metodologie di insegnamento:

Classes Seminars Workshops Preassigned readings.

Contenuti:

All the info about the course are on Moodle - Introduction to Law and Legal Studies - Introduction to the EU Law - Introduction to the EU GDPR - The concept of data; personal, sensitive and economic data; big data - Property of data, choices in the management of data - The right to be forgotten - Civil and criminal aspects of profiling activity - Automatic data processing, human responsibilities - The Data Protection Officer and DP Authorities - Civil and criminal protection of privacy - Sanctioning powers and system - Open Data for the public interest - Big data (collection, analysis, processing) and their influence on fundamental rights - Digital Surveillance - Facial Recognition: Open Issues - Disinformation - Artificial Intelligence in the EU law

Modalità di esame:

Written Exam

Criteri di valutazione:

The grading scale used to assess the students is the Italian one, with the highest score of 30/30 and a minimum score of 18/30 (sufficient). The students will be graded according to their level of theoretical and practical knowledge of the fields covered throughout the course and their capacity to critically reflect on the most contentious legal issues on data-intensive technologies.

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

The course has no official textbooks. Students can study on their notes and the additional material provided by the instructor. Nevertheless, here are some helpful handbooks to approach some modules. These books are not mandatory, and students have sole discretion to refer them. ? Mireille Hildebrandt (2020). Law for Computer Scientists and Other Folk, OUP (available: open access) – especially chapters 2-3-4-5-9-10 ? Paul Voigt, Axel von dem Bussche (2017). The EU General Data Protection Regulation (GDPR). A practical guide, Springer (available through Galileo: unipd libraries) ? European Fundamental Rights Agency (2018). Handbook on European data protection law, Luxembourg (available: open access) ? Karen Yeung, Martin Lodge (2019). Algorithmic Regulation, OUP (Public Law Dept. Library) – especially chapters 2-3-4-6-7-11

MACHINE LEARNING

Titolare: Dott. DAVIDE RIGONI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Sono richieste conoscenze di base di Analisi Matematica, Probabilità e Statistica, Algebra Lineare, e di Programmazione.

Conoscenze e abilità da acquisire:

Il corso presenta i concetti e gli algoritmi fondamentali dell'Apprendimento Automatico, cioè quella classe di tecniche ed algoritmi che a partire dai dati consentono di acquisire nuova conoscenza. Questi sono particolarmente utili per tutti quei problemi per cui è impossibile, o molto difficile, pervenire ad una formalizzazione utilizzabile per la definizione di una soluzione algoritmica ad hoc. Esempi di tali problemi sono i compiti percettivi, come il riconoscimento visivo o il riconoscimento del parlato, e problemi in cui i dati sono affetti da rumore o sono incompleti. Sono previste varie esercitazioni in laboratorio informatico, al fine di consentire allo studente di sperimentare le conoscenze acquisite mediante la loro applicazione a piccoli esempi pratici. Il corso insegna inoltre le competenze necessarie per applicare e valutare modelli di apprendimento automatico in contesti applicativi reali.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso consiste in lezioni e esercizi in laboratorio informatico. Gli esercizi in laboratorio informatico consentono agli studenti di sperimentare, in diversi scenari operativi, le tecniche introdotte a lezione. Gli studenti possono così verificare sperimentalmente i concetti appresi in classe, ed acquisire la capacità di applicarli criticamente di fronte ad un problema concreto.

Contenuti:

Il corso tratta gli argomenti elencati di seguito: - Introduzione Intelligenza Artificiale ed Apprendimento Automatico, quando e perché utilizzare tecniche di Apprendimento Automatico; I diversi paradigmi di Apprendimento Automatico, caratteristiche ed applicazioni; Gli ingredienti Fondamentali dell'Apprendimento Automatico. - Apprendimento Supervisionato: le fondamenta ed i modelli base Modelli di Regressione Lineare; spazio delle ipotesi, rappresentazione e definizione della funzione di costo; ottimizzazione e discesa del gradiente. Modelli di Classificazione Lineare; Regressione Logistica; Regolarizzazione e model selection. - Bontà di un modello e sua valutazione Bias-Variance Tradeoff e come affrontare problemi di overfitting ed underfitting; minimizzazione del rischio e teoria dell'apprendimento; misure di valutazione delle performance, esempi ed applicazioni; come effettuare debugging e

diagnosticare il comportamento di un sistema di Apprendimento Automatico. - Apprendimento Supervisionato: reti neurali e modelli avanzati Reti Neurali Artificiali; perceptron, reti multistrato e cenni di deep learning; apprendimento dei parametri in reti neurali, backpropagation e discesa di gradiente. Support Vector Machines; metodi kernel e classificazione non lineare. Paradigmi "alternativi": metodi non-parametrici (k-NN) e loro applicazioni; alberi di decisione e random forest. - Apprendimento non Supervisionato Clustering: KMeans e cenni agli approcci principali; PCA e riduzione di dimensionalità. - Introduzione ai sistemi di Raccomandazione e alla loro applicazione

Modalità di esame:

Lo studente deve sviluppare, in accordo con il docente, un piccolo progetto applicativo. Inoltre, lo studente deve presentare una relazione scritta sul progetto svolto, in cui si discutono criticamente tutte le questioni trattate durante la sua realizzazione. L'esame consisterà in una breve presentazione e discussione del progetto svolto, ed un esame orale delle tematiche trattate a lezione.

Criteri di valutazione:

Lo scopo del progetto è quello di verificare se lo studente ha acquisito le competenze necessarie ad applicare algoritmi di apprendimento automatico a problemi reali. La discussione critica di quanto realizzato e dei risultati ottenuti costituiscono elemento fondamentale di valutazione. Nel corso dell'esame orale le domande del docente sono finalizzate a valutare il livello di apprendimento delle nozioni impartite nel corso e la capacità dello studente nell'analizzarle criticamente.

Testi di riferimento:

Bishop, Christopher M., Pattern recognition and machine learning. New York: Springer, 2006

Eventuali indicazioni sui materiali di studio:

Le presentazioni mostrate durante le lezioni e link di approfondimento sono resi disponibili su Moodle come materiale di riferimento.

NATURAL LANGUAGE PROCESSING

Titolare: Prof. GIOVANNI DA SAN MARTINO

Mutuato da: Laurea magistrale in Computer Science (Ord. 2021)

Periodo: I anno, 2 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

Tipologie didattiche: 48A; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

OPTIMIZATION FOR DATA SCIENCE

Titolare: Prof. FRANCESCO RINALDI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 2 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

Tipologie didattiche: 48A; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

PROCESS-ORIENTED DATA SCIENCE

Titolare: da definire

Periodo: Il anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

Tipologie didattiche: 27A+27L; 6,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

SEMANTIC DATA MANAGEMENT

Titolare: da definire

Periodo: Il anno, 2 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

Tipologie didattiche: 27A+27L; 6,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

STATISTICAL LEARNING (C.I.)

Titolare: Prof. ALBERTO ROVERATO

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Moduli del C.I.:
Statistical Learning 1 (Mod. A)
Statistical Learning 2 (Mod. B)

STATISTICAL LEARNING 1 (MOD. A)

Titolare: Prof. ALBERTO ROVERATO

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

Tipologie didattiche: 48A; 6,00

Contenuti:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

STATISTICAL LEARNING 2 (MOD. B)

Titolare: Prof. ALBERTO ROVERATO

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 2 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

Tipologie didattiche: 48A; 6,00

Contenuti:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

STOCHASTIC METHODS

Titolare: Prof. MARCO FERRANTE

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Nozioni di base di calcolo differenziale e integrale, algebra lineare e calcolo delle probabilità.

Conoscenze e abilità da acquisire:

Lo scopo del corso è di introdurre metodi e concetti di Calcolo delle Probabilità che hanno un forte impatto come strumenti algoritmici, computazionali e nello studio delle reti.

Attività di apprendimento previste e metodologie di insegnamento:

Lezioni frontali. Esercitazioni in aula

Contenuti:

1. Richiami di Calcolo delle Probabilità. • distribuzioni discrete e continue • variabili aleatorie, valor atteso e valor atteso condizionale • Legge dei grandi numeri e Chernoff Bounds • approssimazione di distribuzioni di probabilità. 2. Catene di Markov e passeggiate aleatorie • Catene di Markov a tempo discreto e relative distribuzioni stazionarie • Monte Carlo (MCMC), convergenza di algoritmi MCMC-based. 3. Variabili Gaussiane di grandi dimensioni • Gaussian Annulus Theorem • Quasi ortogonalità per variabili indipendenti 4. Introduzione ai Grafi Aleatori

Modalità di esame:

Esame scritto

Criteri di valutazione:

Il voto finale è basato sul risultato della prova scritta, il cui scopo principale è verificare la capacità di usare in modo corretto ed efficiente le tecniche esposte, applicandole a problemi concreti.

Testi di riferimento:

Paolo Dai Pra, Stochastic Methods for Data Science. : ,

Eventuali indicazioni sui materiali di studio:

Il docente fornirà delle dispense ed esercizi

TECHNIQUES AND METHODOLOGY OF INNOVATION AND RESEARCH

Titolare: da definire

Periodo: Il anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Italy to Spain)

Tipologie didattiche: 54A; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

Curriculum: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

ALGORITHMS FOR DATA MINING

Titolare: da definire

Periodo: I anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 18A+36L; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

ALGORITHMS, DATA STRUCTURES AND DATABASES

Titolare: da definire

Periodo: I anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 13A+37L; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

BIG DATA MANAGEMENT

Titolare: da definire

Periodo: I anno, 2 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 27A+27L; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

BIOINFORMATICS**Titolare:** Prof.ssa LAURA TREU**Mutuato da:** Laurea magistrale in Computer Science (Ord. 2021)**Periodo:** Il anno, 1 semestre**Indirizzo formativo:** TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)**Tipologie didattiche:** 40A+8E; 6,00**Prerequisiti:**

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

BUSINESS ECONOMIC AND FINANCIAL DATA**Titolare:** Prof.ssa MARIANGELA GUIDOLIN**Mutuato da:** Laurea magistrale in Data Science (Ord. 2023)**Periodo:** Il anno, 1 semestre**Indirizzo formativo:** TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)**Tipologie didattiche:** 48A; 6,00**Prerequisiti:**

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

CLOUD COMPUTING AND DATA ANALYTICS

Titolare: da definire

Periodo: I anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 27A+27L; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

COGNITION AND COMPUTATION

Titolare: Prof. MARCO ZORZI

Mutuato da: Laurea magistrale in Cybersecurity (Ord. 2020)

Periodo: Il anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 48A; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

DATA WAREHOUSING

Titolare: da definire

Periodo: I anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 27A+27L; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

FINAL EXAMINATION

Titolare: da definire

Periodo: Il anno, annuale

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 15A; 15,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

HUMAN DATA ANALYTICS

Titolare: Prof. MICHELE ROSSI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 48A; 6,00

INTERDISCIPLINARY INNOVATION PROJECT

Titolare: da definire

Periodo: I anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 15A+27L; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

INTERNSHIP

Titolare: Prof. TOMASO ERSEGHE

Periodo: Il anno, annuale

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: ; 15,00

Contenuti:

Durante il corso di laurea magistrale in Data Science, gli studenti svolgono uno stage presso un'azienda, un'industria, un centro di ricerca o un laboratorio universitario. Il tirocinio permette di sviluppare nuove competenze, entrare in contatto con l'ambiente di lavoro quotidiano, sviluppare la consapevolezza delle esigenze del mercato e dei clienti. Queste abilità sono molto apprezzate dal mercato del lavoro, rafforzano il curriculum degli studenti e migliorano le loro opportunità di lavoro. Siamo collegati con una serie di selezionati partner industriali e di ricerca, in Italia e nel mondo. Il tirocinio si svolge solitamente durante il 2° anno, 2° semestre e, seppur non obbligatorio, è spesso abbinato alla tesi di laurea magistrale. Una breve guida agli stage è disponibile qui: https://datascience.math.unipd.it/files/a-quick-guide-to-internships-@-data-science_final.pdf Una guida dettagliata per gli studenti iscritti è disponibile nella piattaforma elearning: <https://elearning.unipd.it/math/course/view.php?id=659> Per ulteriori richieste contattaci a stage.datascience@math.unipd.it

Testi di riferimento:

CONTENUTO NON PRESENTE

LAW AND DATA

Titolare: Dott.ssa FIORELLA DAL MONTE

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 48A; 6,00

Prerequisiti:

No prerequisites

Conoscenze e abilità da acquisire:

The course aims to introduce non-law students to a proper understanding of the main legal issues related to the processing of data, personal and non. The first part of the course aims to enable students to approach EU personal data protection regulation. In the second part, instead, students will reflect on the main problems related to the use of data-intensive technologies (big data and artificial intelligence) and the technical and legal solutions now debated.

Attività di apprendimento previste e metodologie di insegnamento:

Classes Seminars Workshops Preassigned readings.

Contenuti:

All the info about the course are on Moodle - Introduction to Law and Legal Studies - Introduction to the EU Law - Introduction to the EU GDPR - The concept of data; personal, sensitive and economic data; big data - Property of data, choices in the management of data - The right to be forgotten - Civil and criminal aspects of profiling activity - Automatic data processing, human responsibilities - The Data Protection Officer and DP Authorities - Civil and criminal protection of privacy - Sanctioning powers and system - Open Data for the public interest - Big data (collection, analysis, processing) and their influence on fundamental rights - Digital Surveillance - Facial Recognition: Open Issues - Disinformation - Artificial Intelligence in the EU law

Modalità di esame:

Written Exam

Criteri di valutazione:

The grading scale used to assess the students is the Italian one, with the highest score of 30/30 and a minimum score of 18/30 (sufficient) (info: here). The students will be graded according to their level of theoretical and practical knowledge of the fields covered throughout the course and their capacity to critically reflect on the most contentious legal issues on data-intensive technologies.

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

The course has no official textbooks. Students can study on their notes and the additional material provided by the instructor. Nevertheless, here are some helpful handbooks to approach some modules. These books are not mandatory, and students have sole discretion to refer them. ? Mireille Hildebrandt (2020). Law for Computer Scientists and Other Folk, OUP (open access: here) – especially chapters 2-3-4-5-9-10 ? Paul Voigt, Axel von dem Bussche (2017). The EU General Data Protection Regulation (GDPR). A practical guide, Springer (unipd access: here) ? European Fundamental Rights Agency (2018). Handbook on European data protection law, Luxembourg (open access: here) ? Karen Yeung, Martin Lodge (2019). Algorithmic Regulation, OUP (Public Law Dept. Library) – especially chapters 2-3-4-6-7-11

MACHINE LEARNING

Titolare: Dott. DAVIDE RIGONI

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: I anno, 2 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 27A+27L; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

MINING UNSTRUCTURED DATA

Titolare: da definire

Periodo: I anno, 2 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 24A+15L; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

MULTIVARIATE ANALYSIS

Titolare: da definire

Periodo: I anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 26A+26L; 6,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

NETWORK SCIENCE

Titolare: Prof. TOMASO ERSEGHE

Mutuato da: Scuola Galileiana di Studi Superiori - Classe di Scienze Naturali

Periodo: Il anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 48A; 6,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

PROCESS-ORIENTED DATA SCIENCE

Titolare: da definire

Periodo: I anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 27A+27L; 6,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

SEMANTIC DATA MANAGEMENT

Titolare: da definire

Periodo: I anno, 2 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 27A+27L; 6,00

Prerequisiti:
CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:
CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:
CONTENUTO NON PRESENTE

Contenuti:
CONTENUTO NON PRESENTE

Modalità di esame:
CONTENUTO NON PRESENTE

Criteri di valutazione:
CONTENUTO NON PRESENTE

Testi di riferimento:
CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:
CONTENUTO NON PRESENTE

STATISTICAL INFERENCE AND MODELING

Titolare: da definire

Periodo: I anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 28A+25L; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

STOCHASTIC METHODS

Titolare: Prof. MARCO FERRANTE

Mutuato da: Laurea magistrale in Data Science (Ord. 2023)

Periodo: Il anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 48A; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

TECHNIQUES AND METHODOLOGY OF INNOVATION AND RESEARCH

Titolare: da definire

Periodo: I anno, 1 semestre

Indirizzo formativo: TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)

Tipologie didattiche: 54A; 6,00

Prerequisiti:

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

VIABILITY OF BUSINESS PROJECTS**Titolare:** da definire**Periodo:** I anno, 1 semestre**Indirizzo formativo:** TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)**Tipologie didattiche:** 18A+36L; 6,00**Prerequisiti:**

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE

VISION AND COGNITIVE SYSTEMS**Titolare:** Prof. LAMBERTO BALLAN**Mutuato da:** Laurea magistrale in Computer Science (Ord. 2021)**Periodo:** II anno, 1 semestre**Indirizzo formativo:** TMDS - Trans Mediterranean Data Science (Study Path Spain to Italy)**Tipologie didattiche:** 32A+16L; 6,00**Prerequisiti:**

CONTENUTO NON PRESENTE

Conoscenze e abilità da acquisire:

CONTENUTO NON PRESENTE

Attività di apprendimento previste e metodologie di insegnamento:

CONTENUTO NON PRESENTE

Contenuti:

CONTENUTO NON PRESENTE

Modalità di esame:

CONTENUTO NON PRESENTE

Criteri di valutazione:

CONTENUTO NON PRESENTE

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

CONTENUTO NON PRESENTE