



Bollettino Notiziario - A.A. 2021/2022

LAUREA MAGISTRALE IN DATA SCIENCE (ORD. 2017)

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

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

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

HUMANITIES: SOCIAL AND ETHICAL IMPACT 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

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

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

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

FINAL EXAMINATION

Titolare: da definire

Periodo: Il anno, annuale

Indirizzo formativo: Corsi comuni

Tipologie didattiche: 15A; 15,00

Contenuti:

The final exam consists in the presentation of the Master Thesis to the Committee of the Final Exam. The thesis (15 CFU) can be based on an internship (15 CFU) that may take place in a private company, a public institution or the University of Padova. The thesis is to be prepared under the guidance of a representative of the hosting institution (co-supervisor) and a Professor or Researcher of the University of Padova (supervisor).

Testi di riferimento:

CONTENUTO NON PRESENTE

STAGE

Titolare: Prof. TOMASO ERSEGHE

Periodo: Il anno, 1 semestre

Indirizzo formativo: Corsi comuni

Tipologie didattiche: 15A; 15,00

Contenuti:

Gli studenti di Data Science hanno la possibilità di svolgere un'attività di tirocinio presso una realtà aziendale o un centro di ricerca (inclusa la stessa Università di Padova), solitamente svolta a valle del primo anno della Laurea Magistrale (LM). Il tirocinio è un periodo di formazione che consente allo studente di acquisire conoscenze e competenze professionali spendibili nel mondo del lavoro, sperimentando una situazione concreta di vita aziendale e sviluppando sensibilità verso le esigenze del mercato e dei clienti. Tipicamente, il tirocinio è perseguito nel secondo semestre del secondo anno di corso e ha una durata tipica che va dai sei agli otto mesi. La LM in Data Science è in contatto diretto con un numero di aziende e centri di ricerca che collaborano con i Professori del Corso. Lo studente può scegliere l'attività di tirocinio sia presso una di queste realtà, sia presso una qualsiasi azienda che operi nel campo dell'analisi dei dati, a patto che questa sia registrata nel database dell'Università degli Studi di Padova (UNIPD). Nel caso non lo fosse, è sempre possibile aggiungere un nuovo partner seguendo una procedura snella e veloce, che va espletata in concerto con il servizio Stage e Tirocini di UNIPD (<https://www.unipd.it/stage>). L'attività da svolgere durante il periodo di tirocinio dovrebbe essere correlata alle tematiche sviluppate nel Corso, come statistical data analysis, machine learning, big data processing, etc. Per ogni tirocinio, devono essere individuate due referenti: 1) il responsabile presso l'azienda, o centro ricerca dove lo studente effettuerà il tirocinio, 2) il Professore responsabile presso l'Università degli Studi di Padova, che affiancherà l'attività di tirocinio fornendo il supporto tecnico necessario, garantendo la qualità tecnica del periodo formativo e assicurandosi della coerenza con le tematiche trattate nella LM in Data Science. Spesso, anche se non obbligatoriamente, l'attività svolta durante il tirocinio diventa oggetto della Tesi di Laurea Magistrale. Questo dipende dalla mutua soddisfazione tra studente e azienda / centro di ricerca. In sintesi, gli obiettivi formativi per lo studente includono - entrare in contatto con realtà aziendali o di ricerca, respirandone l'ambiente, le modalità di lavoro e le esigenze; - applicare sviluppi teorici e algoritmi da essi derivati a problemi reali, che trovino possibilmente sbocco in un prodotto; - lavorare alla Tesi Magistrale, sviluppando tematiche di interesse per il mondo industriale o della ricerca; - creare sinergie e contatti utili per una possibile e soddisfacente occupazione futura, a valle della Laurea. Per ulteriori informazioni, lo studente interessato a questa attività è può mettersi in contatto con i nostri uffici utilizzando l'email preposta: internships.datascience@math.unipd.it

Testi di riferimento:

CONTENUTO NON PRESENTE

Curriculum: Data Science - Piano di studio Biological Data Analytics

BIG DATA COMPUTING

Titolare: Prof. ANDREA ALBERTO PIETRACAPRINA

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

DEEP LEARNING

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

FUNDAMENTALS OF INFORMATION SYSTEMS

Titolare: Prof. GIORGIO MARIA DI NUNZIO

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: I anno, 1 semestre

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

Tipologie didattiche: 96A; 12,00

Prerequisiti:

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

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:

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: Prof. GIOVANNI DA SAN MARTINO

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

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 vari aspetti di biologia cellulare. Conoscenza di analisi di sistemi dinamici semplici, stocastici, spazio-temporali. Abilità di fare analisi di sistemi dinamici nonlineari e stocastici. 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. Propagazione di potenziale di azione. Dinamica stocastica di canali ionici. Dinamica intracellulare di calcio, onde di calcio. Analisi matematica di sistemi dinamici, stocastici, spazio-temporali. Biforcazioni.

Modalità di esame:

Progetto e esame scritto.

Criteri 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:

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

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 (es., MatLab, Python, C, Java); conoscenze in Analisi Matematica e Algebra Lineare; ogni ulteriore conoscenza dei processi di rete in economia, biologia, telecomunicazioni, semantica, etc. può essere utile.

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. Conoscere i principali ambiti di applicazione, anche interdisciplinari, delle tecniche studiate 7. Essere in grado di sintetizzare l'analisi di una rete in un documento professionale 8. Essere in grado di implementare al calcolatore algoritmi atti all'analisi delle reti

Attività di apprendimento previste e metodologie di insegnamento:

Il modulo prevede: - 16 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; - 4 lezioni interattive finalizzate a progetti interdisciplinari con gli studenti del corso gemello di Social Network Analytics.

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; Evolution networks. 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 di comunità - Dendrogrammi; Metodo di Girvan Newman e betweenness; Ottimizzazione della misura di modularità (Louvain modularity); Clustering spettrale; Consensus clustering; Misura della similarità tra diverse soluzioni; Algoritmi per comunità sovrapposte. 6. Rappresentazione di una rete: funzionalità grafiche di Gephi e R/Python; algoritmi force-directed per la visualizzazione grafica di una rete. 7. Twitter Lab - Come estrarre una rete

semantica da Twitter.

Modalità di esame:

La verifica delle conoscenze e delle abilità attese viene effettuata tramite 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 stesura di una elaborato; 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.

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 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:

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

Eventuali indicazioni sui materiali di studio:

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

OMICS IN HUMAN DISEASE

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 dell'invecchiamento. Inoltre vengono descritte le principali tecnologie disponibili oggi per lo studio multi-dimensionale e su larga scala dei fenomeni biologici con particolare attenzione ai dati prodotti dalle tecnologie omiche in ambito clinico e non, e alla loro analisi. Lo studente acquisirà gli strumenti bioinformatici per analizzare i dati omici e formulare ipotesi biologiche riguardo ai meccanismi molecolari alla base di malattie complesse, l'interpretazione delle mutazioni genetiche e l'identificazione di possibili target farmacologici

Attività di apprendimento previste e metodologie di insegnamento:

Il corso si compone di lezioni frontali, esercitazioni pratiche al computer, sviluppo di un progetto di gruppo e presentazione dello stesso con discussione critica. Le esercitazioni forniscono gli strumenti computazionali per svolgere un'analisi bioinformatica riguardante l'interpretazioni dei dati omici e clinici relativi ad una malattia complessa. La presentazione del progetto permette di valutare l'effettiva comprensione del problema e i punti di forza e debolezze dell'analisi proposta.

Contenuti:

Il corso si compone di due parti: 1. Introduzione alla fisiologia umana e alle malattie complesse (4 CFU): a. Cenni di biochimica e biologia cellulare b. Introduzione alla fisiologia umana c. Ereditarietà e patogenicità delle mutazioni d. Meccanismi molecolari delle patologie complesse: cancro, neurosviluppo e invecchiamento 2. Biochimica computazionale (2 CFU): a. Banche dati biologiche b. Metodi per l'analisi di network di interazione molecolare c. Metodi per l'analisi di dati di sequenziamento d. Metodi per l'analisi di dati clinici e. Metodi per l'analisi e l'interpretazione di varianti patogenetiche

Modalità di esame:

L'esame si compone in due parti: Una valutazione delle conoscenze biologiche riguardo la fisiologia umana e la biologia cellulare relative alla malattie complesse. Una valutazione delle conoscenze pratiche acquisite durante le esercitazioni, tramite la valutazione del progetto riguardante l'analisi computazionale di una malattia complessa.

Criteri di valutazione:

Viene valutata: La comprensione di concetti presentati a lezione La capacità di applicare le nozioni fornite a lezione su problemi reali La capacità critica di saper progettare un'analisi computazionale 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 (appena disponibili), la letteratura di riferimento e il materiale delle esercitazioni (scripts e tutorial).

OPTIMIZATION FOR DATA SCIENCE

Titolare: Prof. FRANCESCO RINALDI

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: I anno, 2 semestre

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

Tipologie didattiche: 48A; 6,00

STATISTICAL LEARNING

Titolare: Prof. ALBERTO ROVERATO

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

Prerequisiti:

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

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. 2017)

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. 2017)

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. 2017)

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:

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

SYSTEMS BIOLOGY

Titolare: Prof. GABRIELE SALES

Mutuato da: Laurea magistrale in Molecular Biology (Ord. 2020)

Periodo: I anno, 1 semestre

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

Tipologie didattiche: 40A+16E; 3,00

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

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:

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

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:
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

FINANCIAL MATHEMATICS FOR DATA SCIENCE

Titolare: Prof. MARTINO GRASELLI

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: 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:

FUNDAMENTALS OF INFORMATION SYSTEMS
--

Titolare: Prof. GIORGIO MARIA DI NUNZIO

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

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 and (group) project. The project is due by the end of the first session of exams.

Criteri di valutazione:

The written exam and the project 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. The final grade is obtained as the weighted sum of the grades of the written exam (80%) and the project (20%).

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/> and <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

HUMAN COMPUTER INTERACTION

Titolare: Prof. LUCIANO GAMBERINI

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

Periodo: I anno, 1 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:

NON FREQUENTANTI: L'esame sarà orale con 3 domande sul libro, una delle quali proposta come esercizio (vedere il libro per trovare esempi). In caso di permanenza della situazione di emergenza l'orale sarà svolto on-line FREQUENTANTI: L'esame sarà basato su un lavoro personale di ricerca da svolgere durante il corso come lavoro per casa e su una breve prova orale di presentazione e discussione del medesimo. Un report sintetizzerà il lavoro svolto in modo simile a un paper scientifico o a un report professionale. La ricerca da sviluppare verrà riassunta in un report che includerà i seguenti punti: 1 - introduzione, contestualizzazione teorica, benchmarking 2 - processo di design/co-design 3 - sviluppo/modifica prototipo (per chi ha meno esperienza tecnica ci sono soluzioni) 4 - valutazione - analisi dei dati (es: UX, usability, presence, acceptance) 5 - risultati - discussione finale Qualsiasi tecnologia potrà essere sviluppata/adottata per la ricerca purché sia interattiva (es: web, robot, realtà virtuale/aumentata, dispositivi smart home, strumenti di lavoro, tools di Arduino, veicoli, strumenti musicali, apps. su cellulare/Tablet).

Criteri di valutazione:

NON PARTECIPANTI: Ogni domanda ha lo stesso valore, in altri termini ogni domanda ha un peso di 10 su 30. Il punteggio finale sarà la somma del punteggio di ogni domanda. PARTECIPARE La valutazione si baserà in gran parte sulla qualità del report e delle sue parti con un punteggio di 5 punti per ciascuna di esse, per un punteggio massimo di 25. I restanti 5 punti saranno basati sulla presentazione orale della ricerca.

Testi di riferimento:

Helen Sharp, Jenny Preece, Yvonne Rogers, Interaction Design: Beyond Human-computer Interaction. : Wiley, 2019

Eventuali indicazioni sui materiali di studio:

FREQUENTANTI INTERACTION DESIGN (5th 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 (5th ed.)- di Helen Sharp, Jenny Preece e Yvonne Rogers è l'unico materiale da utilizzare per questo corso se non lo si frequenta.

HUMAN DATA ANALYTICS

Titolare: Prof. MICHELE ROSSI

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: I anno, 1 semestre

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

Tipologie didattiche: 48A; 6,00

KNOWLEDGE AND DATA MINING

Titolare: da definire

Periodo: I anno, 2 semestre

Indirizzo formativo: Data Science - Piano di studio Cognitive, Social and Economic 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: Prof. GIOVANNI DA SAN MARTINO

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: I anno, 1 semestre

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

OPTIMIZATION FOR DATA SCIENCE

Titolare: Prof. FRANCESCO RINALDI

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: I anno, 2 semestre

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

PROCESS MINING

Titolare: Prof. MASSIMILIANO DE LEONI

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

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"

Conoscenze e abilità da acquisire:

Lo scopo del corso e' quello di introdurre e studiare metodi e concetti che riguardano la modellazione e l'analisi di processi di business. In dettaglio, il corso inizia con illustrare i principi di modellazione di processi come ci si attende che vengano eseguiti (i cosiddetti "TO-BE processes"), usando i più comuni linguaggi di modellazione (BPMN, Petri Nets e Declare), e con descrivere le tecniche di analisi della loro correttezza. Successivamente, il corso si focalizza sul Process Mining, che si pone l'obiettivo di utilizzare lo storico dei dati transazionali (i log degli eventi) per analizzare e migliorare i processi come sono eseguiti (noti come "AS-IS processes"), che possono differire da quelli attesi. Un' enfasi speciale verrà messa sugli algoritmi e sulla loro implementazione in piattaforme software open-source e commerciali. Sessioni pratiche complementano il corso affinché gli studenti possano toccare con mano i problemi di modellazione e mining di processi aziendali in contesti reali/realistici, utilizzando dati reali.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso consiste di lezioni, esercitazioni e sessioni di laboratorio. Le sessioni di laboratorio creeranno l'opportunità di "sporcarsi le mani" su dati reali, così da comprendere la complessità di eseguire un progetto di Process Mining e di lavorare con dati di processo reali.

Contenuti:

Il corso affronterà i seguenti argomenti: 1. MODELLAZIONE ED ANALISI ATTRAVERSO IL LINGUAGGIO BPMN - Modellazione di processo base ed avanzata in BPMN - Analisi Qualitativa - Analisi Quantitativa 2. MODELLAZIONE ED ANALISI 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 Petri Nets: Principi e Verifica. 3. 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 Inductive 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 Sociali 4. TECNICHE PREDITTIVE E PROCESSI DI BUSINESS - Tecniche di Base di Predictive Process Monitoring - Tecniche Avanzate di Predictive Process Monitoring

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. : Springer, 2013 Wil M.P. van der Aalst and Christian Stahl, Modeling Business Processes: A Petri Net-Oriented Approach.. : Information Systems, 2011 Wil M.P. van der Aalst, Process Mining: Data Science in Action. Berlin: Springer-Verlag, 2016

Eventuali indicazioni sui materiali di studio:

Slides, Eserciziari, Esercitazioni di Laboratori Autoguidate e Articoli Scientifici.

STATISTICAL LEARNING

Titolare: Prof. ALBERTO ROVERATO

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

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. 2017)

Periodo: I anno, 1 semestre

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

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. 2017)

Periodo: I anno, 2 semestre

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

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. 2017)

Periodo: I anno, 1 semestre

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

Curriculum: Data Science - Piano di studio Machine Learning for Intelligence systems

BIG DATA COMPUTING

Titolare: Prof. ANDREA ALBERTO PIETRACAPRINA

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 Intelligence systems

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Il corso ha i seguenti prerequisiti: competenze relative al progetto e all'analisi di algoritmi e strutture dati, conoscenza delle nozioni fondamentali di calcolo delle probabilità e statistica, e capacità di programmazione in Java o Python.

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. Programming frameworks: MapReduce, Apache Spark Reducing input size (Case study: clustering) Reducing output size (Case study: frequent itemsets) Streaming framework.

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:

J. Leskovec, A. Rajaraman and J. Ullman, Mining Massive Datasets. : Cambridge University Press, 2014

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. GIORGIO VALLE

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

Periodo: Il anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligence 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 l'analisi del trascrittoma, la predizione e annotazione genica, la ricerca di pattern e motivi per la predizione delle strutture proteiche. Inoltre viene discusso il ruolo della bioinformatica nell'analisi di genomi individuali e nella medicina personalizzata.

Modalità di esame:

L'esame si articola in tre parti: 1) una sessione pratica nella quale lo studente deve descrivere un progetto di analisi di dati che deve essere consegnato almeno due giorni prima della data dell'esame, 2) una sessione di quiz su Moodle, che si svolgerà all'inizio dell'appello d'esame e 3) una discussione orale in cui lo studente deve descrivere il progetto che ha realizzato e rispondere a domande sui contenuti del corso. Un continuo monitoraggio sarà attuato durante l'intera durata del corso per verificare la comprensione degli studenti.

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. SILVIO TOSATTO

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Data Science - Piano di studio Machine Learning for Intelligence 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 quattro parti, che corrispondono ai diversi tipi di dati biologici: 1) Sequenze 1.1) DNA e proteine 1.2) Banche dati 1.3) Allineamenti 2) Strutture 2.1) Folding proteico 2.2) Banche dati 2.3) Predizione di struttura 3) Letteratura 3.1) Articoli scientifici 3.2) Banche dati 3.3) Text mining 3.4) Funzione 4) Reti di interazione 4.1) Regioni non-globulari 4.2) Interazioni biologiche 4.3) Banche dati 4.4) Proprieta' emergenti

Modalità di esame:

L'esame si compone di tre parti separate, che devono essere superate tutte: (i valori tra parentesi indicano i pesi per il voto complessivo) 1) Compitino di meta' semestre (ca. 17%) Domande a risposta aperta 2) Progetto (ca. 50%) Programmazione e analisi dataset biologico; relazione sul progetto 3) Esame finale (ca. 33%) Orale, con domande su progetto e corso

Criteri di valutazione:

Viene valutata: 1) la comprensione di concetti e gli algoritmi presentati a lezione 2) la capacità di applicare le nozioni fornite a lezione su problemi reali 3) la capacità critica di saper utilizzare i metodi nei modi più opportuni, scegliendo tra le alternative possibili 4) la capacità di sviluppare software riutilizzabile estendendo librerie esistenti 5) 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 (appena disponibili) e le registrazioni audio (podcast), le dispense e la letteratura usata per i progetti.

BUSINESS ECONOMIC AND FINANCIAL DATA

Titolare: Prof.ssa MARIANGELA GUIDOLIN

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: II anno, 1 semestre

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

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:

Hyndman, R. J., Athanasopoulos, G., Forecasting: principles and practice.. : OTexts., 2018 Azzalini, A., Scarpa, B., Data analysis and data mining: An introduction.. USA: OUP, 2012 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 Machine Learning for Intelligence systems

Tipologie didattiche: 48A; 6,00

FUNDAMENTALS OF INFORMATION SYSTEMS

Titolare: Prof. GIORGIO MARIA DI NUNZIO

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: I anno, 1 semestre

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

Tipologie didattiche: 96A; 12,00

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 Intelligence systems

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

KNOWLEDGE AND DATA MINING

Titolare: Dott. LUCIANO SERAFINI

Periodo: I anno, 2 semestre

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

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

LAW AND DATA

Titolare: Dott.ssa ELISA SPILLER

Periodo: I anno, 1 semestre

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

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 AND DEEP LEARNING (C.I.)

Titolare: Prof. GIOVANNI DA SAN MARTINO

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

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.:
Machine and deep learning (Mod. A)
Machine and deep learning (Mod. B)

MACHINE AND DEEP LEARNING (MOD. A)

Titolare: Prof. GIOVANNI DA SAN MARTINO

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: I anno, 1 semestre

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

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

MACHINE AND DEEP LEARNING (MOD. B)

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 Intelligence systems

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

OPTIMIZATION FOR DATA SCIENCE

Titolare: Prof. FRANCESCO RINALDI

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: I anno, 2 semestre

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

Tipologie didattiche: 48A; 6,00

PROCESS MINING

Titolare: Prof. MASSIMILIANO DE LEONI

Periodo: I anno, 1 semestre

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

Tipologie didattiche: 48A; 6,00

STATISTICAL LEARNING

Titolare: Prof. ALBERTO ROVERATO

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

Prerequisiti:

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

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. 2017)

Periodo: I anno, 1 semestre

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

Tipologie didattiche: 48A; 6,00

STATISTICAL LEARNING 2 (MOD. B)

Titolare: Prof. ALBERTO ROVERATO

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: I anno, 2 semestre

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

Tipologie didattiche: 48A; 6,00

STOCHASTIC METHODS

Titolare: Prof. MARCO FERRANTE

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: I anno, 1 semestre

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

Tipologie didattiche: 48A; 6,00

Curriculum: Data Science - Piano di studio Mathematics of Data Science

FINANCIAL MATHEMATICS FOR DATA SCIENCE

Titolare: Prof. MARTINO GRASELLI

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

Periodo: I anno, 2 semestre

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

Tipologie didattiche: 48A; 6,00

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

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 computer networking. The ability of processing data effectively and efficiently will be gained using Python, which is possibly the reference programming language for data scientists. Ultimately, students will acquire coding skills to collect, clean, visualize, and analyse data, and more generally to tackle with 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. Basics of systems for storage of streams of data are presented as well. The networking submodule provides an introduction to fundamental concepts in the design and implementation of computer communication networks, their protocols, and applications. Topics covered in this part include: layered network architecture, data link protocols, network and transport protocols and applications. Examples will be drawn from the Internet TCP/IP protocol suite. After that, advanced and emerging networking paradigms aimed at addressing QoS and engineering flexibility of current infrastructure networks are introduced. Topics covered range from software defined networking to cloud provisioning schemes and data centers.

Attività di apprendimento previste e metodologie di insegnamento:

The course consists of lectures.

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). Eventually, at the end of this submodule students will be able to implement all the stages of a typical machine learning pipeline: from collecting data to building predictive models for solving either a classification or a regression problem. - Databases This submodule is dedicated to data storage, and it covers the following topics: Architecture of Database management systems (DBMS). Relational modeling. 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. - Networking This submodule allows students to get familiar with computer networking. In particular, it focuses on the following topics: Networking Fundamentals: Network architectures (OSI Model); TCP and UDP Transport layer protocols; IP Addressing and Routing; Link Layer Forwarding; DNS and DHCP. Advanced Networking: Virtual LAN (VLAN) and Virtual eXtensible Lan (VXLAN), Software Defined Networking: control, data plane and virtualization; concepts on Cloud Computing: service and deployment models: data centers architectures, topologies, addressing, routing, traffic characteristics; Case Study: The Web of Things (IoT standards and protocols).

Modalità di esame:

The student is expected to pass a written and an oral exam.

Criteri di valutazione:

The written and the oral exams 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/>

GAME THEORY**Titolare:** ELVINA GINDULLINA**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 anche basilare di teoria della probabilità.

Conoscenze e abilità da acquisire:

L'insegnamento prevede l'acquisizione delle seguenti conoscenze e abilità, suddivise in due insiemi. 1: parte base. 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. 2: parte applicativa. Sapere applicare i concetti della teoria dei giochi a scenari pratici, specialmente di tipo ICT; in questo contesto, e' di particolare interesse l'abilità di contestualizzare la teoria dei giochi come strumento di valutazione per l'efficacia della risoluzione tramite procedure multi-agente distributed.

Attività di apprendimento previste e metodologie di insegnamento:

Lezioni convenzionali con il supporto di slide. Prevista interazione su piattaforma moodle. Caricamento video via Kaltura.

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 Giochi cooperativi, il nucleo, il valore di Shapley Allocazione delle risorse Utilità, scelte e paradossi Giochi potenziali, coordinazione Algoritmi bio-inspired Giochi evolutivi Reti cognitive Selfish routing Sistemi multi-input con teoria dei giochi

Modalità di esame:

In qualunque caso l'esame comprende un test scritto obbligatorio a libro aperto, dove vengono sottoposti diversi problemi di game theory allo studente su argomenti toccati durante il corso. Per ogni esercizio, vengono poste più domande, tipicamente tre. Per frequentanti, l'esame può coinvolgere lo sviluppo di un progetto in gruppi di 1-3 persone, su argomenti del corso applicati alle ICT. L'adesione a questa modalità e l'argomento del progetto sono concordati con il docente durante il corso. Se il test scritto è sufficiente, si può registrare il voto conseguito come voto finale dell'esame. Si può ulteriormente discutere il progetto sviluppato durante il corso con un esame orale, da svolgersi dopo l'esame scritto. Questi esami orali si svolgono nella stessa giornata di un esame scritto, ma non necessariamente bisogna presentarsi nella stessa giornata per l'esame scritto e la discussione orale del progetto. La discussione orale integra il voto dello scritto.

Criteri di valutazione:

Ogni domanda nei test scritti viene valutata fino a un massimo di 3 punti. La discussione del progetto viene valutata fino a 10 punti. Il voto finale è la somma numerica dei punteggi individuali delle domande e della discussione del progetto (se presente), limitata a 30. Un punteggio di 30 e lode è assegnato agli studenti il cui punteggio numerico è superiore a 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 sia con le tematiche del corso che con le metodologie ingegneristiche tipiche dell'ICT - la qualità dell'esposizione orale - 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:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

Diversi libri forniscono una trattazione generale di teoria dei giochi. A mero titolo di suggerimento, si può usare il libro di Tadelis come riferimento in senso generale. Questa parte comunque dovrebbe essere integrata con materiale per le applicazioni. Il libro di MacKenzie e DaSilva è un buon esempio, anche se non è obbligatorio usare un libro per questo scopo (si può fare riferimento anche a materiale trovato in rete). In ogni caso, il docente fornirà agli studenti tutte le dispense delle lezioni e appunti aggiuntivi.

HIGH DIMENSIONAL PROBABILITY FOR DATA SCIENCE**Titolare:** Prof. MARCO FORMENTIN**Periodo:** I anno, 1 semestre**Indirizzo formativo:** Data Science - Piano di studio Mathematics of Data Science**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

HUMAN DATA ANALYTICS**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. Si noti che il Professore, effettuerà dei brevi ripassi per le tecniche e teorie di cui sopra, ogni volta lo si renda 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, si consiglia l'esame "Machine Learning" della Laurea Magistrale ICT for Internet and Multimedia, codice: INP6075419.

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 in seno a problemi reali 4. Conoscere i principi e le tecniche dell'apprendimento supervisionato con particolare riferimento alle reti neurali (feed forward e convoluzionali), 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 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 architetture 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) è reso disponibile sul sito personale del docente (protetto via username e password): <http://www.dei.unipd.it/~rossi/courses/HumanData/HDA.html> Le credenziali di accesso verranno comunicate dal docente nella lezione di introduzione al 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 Part II – Tecniche di quantizzazione vettoriale (12 ore) - Quantizzazione Vettoriale (QV): -- Significato, utilità, metriche -- K-means, soft K-means, Expectation Maximization - Algoritmi per la QV non-supervisionata: --- Mappe Auto-Organizzanti (SOM), Reti Neurali "Gas" (GNG) - 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 II – Analisi di dati sequenziali (10 ore) - Modelli di Markov nascosti (HMM): --- Massima verosimiglianza per le HMM --- Algoritmo "Forward-backward" --- Algoritmo di somma-prodotto, algoritmo di Viterbi - Applicazioni --- Autenticazione: identificazione di utenti dalle dinamiche di digitazione su tastiere numeriche --- Riconoscimento vocale: estrazione di feature vocali, riconoscimento vocale tramite modelli di Markov nascosti Part III - Reti Neurali "Deep" (10 ore) - Tecniche di discesa del gradiente e concetti generali (apprendimento supervisionato, modelli di costo, etc.) - Reti neurali "feed forward": modelli, allenamento delle reti, algoritmo di "back-propagation" - Reti neurali convoluzionali (CNN): struttura, blocchi costituenti, allenamento - Applicazioni: apprendimento di attività --- Attività e sensori: definizioni, classi di attività --- Feature: sequence of feature, feature statistiche, feature spettrali, feature correlate al contensto e alle attività --- Riconoscimento di attività: segmentazione, finestre mobili, segmentazione non-supervisionata, misure prestazionali e risultati - Autenticazione utenti da segnali di moto: combinazione di CNN-SVM and teoria della stima sequenziale - Object / face recognition through CNN 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" e quelle convoluzionali, 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 - Soluzione di un problema di inferenza statistica - Reti neurali con struttura "feed forward" - Reti neurali convoluzionali

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 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:

Bishop, Christopher M., Pattern recognition and machine learning Christopher M. Bishop. New York: Springer, 2006 Bengio, Yoshua; Courville, Aaron; Goodfellow, Ian, Deep Learning. Cambridge: MIT Press, 2016 Watt, Jeremy; Borhani, Reza; K. Katsaggelos, Aggelos, Machine learning refined. New York: Cambridge University Press, 2016

Eventuali indicazioni sui materiali di studio:

Report tecnici, articoli scientifici, software e altro materiale vario saranno resi disponibili dal docente qualora lo si renda necessario. Il materiale sarà reso disponibile tramite il sito del corso. Altri libri utili 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 lo studio di modelli audio e l'utilizzo di HMM per il riconoscimento vocale: - D. Jurafsky, J. H. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition," Prentice Hall, 2nd Edition, 2008.

MACHINE AND DEEP LEARNING (C.I.)

Titolare: Prof. GIOVANNI DA SAN MARTINO

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

Prerequisiti:

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

Moduli del C.I.:

Machine and deep learning (Mod. A)

Machine and deep learning (Mod. B)

MACHINE AND DEEP LEARNING (MOD. A)

Titolare: Prof. GIOVANNI DA SAN MARTINO

Periodo: I anno, 1 semestre

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

Tipologie didattiche: 48A; 6,00

MACHINE AND DEEP LEARNING (MOD. B)

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

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:

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: 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

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

STATISTICAL LEARNING

Titolare: Prof. ALBERTO ROVERATO

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

Prerequisiti:

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

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

Contenuti:

Part 1: Modes of Inference - Data: summary statistics, displaying distributions; exploring relationships - Likelihood: the likelihood, likelihood for several parameters - Estimation: maximum likelihood estimation; accuracy of estimation; the sampling distribution of an estimator; the bootstrap - Hypothesis testing - Other approaches to inference

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:

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

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

Contenuti:

Part 2 - Models : normal linear models; inference for linear models; generalized linear models; inference for generalized linear models - Model selection - Multivariate Analysis: dimension reduction; classification; clustering

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, Daniela Witten, Trevor Hastie and Robert Tibshirani, An Introduction to Statistical Learning. : Springer, 2013 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

Periodo: I anno, 1 semestre

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

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 Mathematics of Data Science

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

Curriculum: Piano di studio Biological Data Analytics**BIOINFORMATICS**

Titolare: Prof. GIORGIO VALLE

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

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Biological Data Analytics

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 l'analisi del trascrittoma, la predizione e annotazione genica, la ricerca di pattern e motivi per la predizione delle strutture proteiche. Inoltre viene discusso il ruolo della bioinformatica nell'analisi di genomi individuali e nella medicina personalizzata.

Modalità di esame:

L'esame si articola in tre parti: 1) una sessione pratica nella quale lo studente deve descrivere un progetto di analisi di dati che deve essere consegnato almeno due giorni prima della data dell'esame, 2) una sessione di quiz su Moodle, che si svolgerà all'inizio dell'appello d'esame e 3) una discussione orale in cui lo studente deve descrivere il progetto che ha realizzato e rispondere a domande sui contenuti del corso. Un continuo monitoraggio sarà attuato durante l'intera durata del corso per verificare la comprensione degli studenti.

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. SILVIO TOSATTO

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

COGNITIVE, BEHAVIORAL AND SOCIAL DATA

Titolare: Prof. GIUSEPPE SARTORI

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

GAME THEORY

Titolare: ELVINA GINDULLINA

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

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Un corso anche basilare di teoria della probabilità.

Conoscenze e abilità da acquisire:

L'insegnamento prevede l'acquisizione delle seguenti conoscenze e abilità, suddivise in due insiemi. 1: parte base. 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. 2: parte applicativa. Sapere applicare i concetti della teoria dei giochi a scenari pratici, specialmente di tipo ICT; in questo contesto, e' di particolare interesse l'abilità di contestualizzare la teoria dei giochi come strumento di valutazione per l'efficacia della risoluzione tramite procedure multi-agente distributed.

Attività di apprendimento previste e metodologie di insegnamento:

Lezioni convenzionali con il supporto di slide. Prevista interazione su piattaforma moodle. Caricamento video via Kaltura.

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 Giochi cooperativi, il nucleo, il valore di Shapley Allocazione delle risorse Utilità, scelte e paradossi Giochi potenziali, coordinazione Algoritmi bio-inspired Giochi evolutivi Reti cognitive Selfish routing Sistemi multi-input con teoria dei giochi

Modalità di esame:

In qualunque caso l'esame comprende un test scritto obbligatorio a libro aperto, dove vengono sottoposti diversi problemi di game theory allo studente su argomenti toccati durante il corso. Per ogni esercizio, vengono poste piu' domande, tipicamente tre. Per frequentanti, l'esame puo' coinvolgere lo sviluppo di un progetto in gruppi di 1-3 persone, su argomenti del corso applicati alle ICT. L'adesione a questa modalita' e l'argomento del progetto sono concordati con il docente durante il corso. Se il test scritto e' sufficiente, si puo' registrare il voto conseguito come voto finale dell'esame. Si puo' ulteriormente discutere il progetto sviluppato durante il corso con un esame orale, da svolgersi dopo l'esame scritto. Questi esami orali si svolgono nella stessa giornata di un esame scritto, ma non necessariamente bisogna presentarsi nella stessa giornata per l'esame scritto e la discussione orale del progetto. La

discussione orale integra il voto dello scritto.

Criteri di valutazione:

Ogni domanda nei test scritti viene valutata fino a un massimo di 3 punti. La discussione del progetto viene valutata fino a 10 punti. Il voto finale e' la somma numerica dei punteggi individuali delle domande e della discussione del progetto (se presente), limitata a 30. Un punteggio di 30 e lode e' assegnato agli studenti il cui punteggio numerico e' superiore a 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 capacita' di problem solving - la capacita' di discussione e verifica ex-post della soluzione trovata Nella valutazione del progetto (se presente) vengono tenuti in considerazione: - l'originalita' della proposta e la pertinenza sia con le tematiche del corso che con le metodologie ingegneristiche tipiche dell'ICT - la qualita' dell'esposizione orale - la capacita' di lavoro di gruppo e la presenza di singoli contributi attribuibili ai partecipanti al progetto - la capacita' di trarre conclusioni significative dal punto di vista scientifico grazie alle metodologie apprese nel corso

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

Diversi libri forniscono una trattazione generale di teoria dei giochi. A mero titolo di suggerimento, si può usare il libro di Tadelis come riferimento in senso generale. Questa parte comunque dovrebbe essere integrata con materiale per le applicazioni. Il libro di MacKenzie e DaSilva è un buon esempio, anche se non è obbligatorio usare un libro per questo scopo (si può fare riferimento anche a materiale trovato in rete). In ogni caso, il docente fornirà agli studenti tutte le dispense delle lezioni e appunti aggiuntivi.

HUMAN DATA ANALYTICS

Titolare: Prof. MICHELE ROSSI

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

LAW AND DATA

Titolare: Dott.ssa ELISA SPILLER

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

MATHEMATICAL CELL BIOLOGY

Titolare: da definire

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

Periodo: Il anno, 1 semestre

Indirizzo formativo: 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 vari aspetti di biologia cellulare. Conoscenza di analisi di sistemi dinamici semplici, stocastici, spazio-temporali. Abilità di fare analisi di sistemi dinamici nonlineari e stocastici. 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. Propagazione di potenziale di azione. Dinamica stocastica di canali ionici. Dinamica intracellulare di calcio, onde di calcio. Analisi matematica di sistemi dinamici, stocastici, spazio-temporali. Biforcazioni.

Modalità di esame:

Progetto e esame scritto.

Criteri 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:

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: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Biological Data Analytics

Tipologie didattiche: 48A; 6,00

OMICS IN HUMAN DISEASE

Titolare: Prof. DENIS DOMINIQUE MARTINVALET

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

Indirizzo formativo: 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 dell'invecchiamento. Inoltre vengono descritte le principali tecnologie disponibili oggi per lo studio multidimensionale e su larga scala dei fenomeni biologici con particolare attenzione ai dati prodotti dalle tecnologie omiche in ambito clinico e non, e alla loro analisi. Lo studente acquisirà gli strumenti bioinformatici per analizzare i dati omici e formulare ipotesi biologiche riguardo ai meccanismi molecolari alla base di malattie complesse, l'interpretazione delle mutazioni genetiche e l'identificazione di possibili target farmacologici

Attività di apprendimento previste e metodologie di insegnamento:

Il corso si compone di lezioni frontali, esercitazioni pratiche al computer, sviluppo di un progetto di gruppo e presentazione dello stesso con discussione critica. Le esercitazioni forniscono gli strumenti computazionali per svolgere un'analisi bioinformatica riguardante l'interpretazione dei dati omici e clinici relativi ad una malattia complessa. La presentazione del progetto permette di valutare l'effettiva comprensione del problema e i punti di forza e debolezze dell'analisi proposta.

Contenuti:

Il corso si compone di due parti: 1. Introduzione alla fisiologia umana e alle malattie complesse (4 CFU): a. Cenni di biochimica e biologia cellulare b. Introduzione alla fisiologia umana c. Ereditarietà e patogenicità delle mutazioni d. Meccanismi molecolari delle patologie complesse: cancro, neurosviluppo e invecchiamento 2. Biochimica computazionale (2 CFU): a. Banche dati biologiche b. Metodi per l'analisi di network di interazione molecolare c. Metodi per l'analisi di dati di sequenziamento d. Metodi per l'analisi di dati clinici e. Metodi per l'analisi e l'interpretazione di varianti patogenetiche

Modalità di esame:

L'esame si compone in due parti: Una valutazione delle conoscenze biologiche riguardo la fisiologia umana e la biologia cellulare relative alla malattie complesse. Una valutazione delle conoscenze pratiche acquisite durante le esercitazioni, tramite la valutazione del progetto riguardante l'analisi computazionale di una malattia complessa.

Criteri di valutazione:

Viene valutata: La comprensione di concetti presentati a lezione La capacità di applicare le nozioni fornite a lezione su problemi reali La capacità critica di saper progettare un'analisi computazionale 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 (appena disponibili), la letteratura di riferimento e il materiale delle esercitazioni (scripts e tutorial).

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

BUSINESS ECONOMIC AND FINANCIAL DATA

Titolare: Prof.ssa MARIANGELA GUIDOLIN

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

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

Tipologie didattiche: 48A; 6,00

COGNITION AND COMPUTATION

Titolare: Prof. MARCO ZORZI

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

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

Tipologie didattiche: 48A; 6,00

COGNITIVE, BEHAVIORAL AND SOCIAL DATA

Titolare: Prof. GIUSEPPE SARTORI

Periodo: Il anno, 1 semestre

Indirizzo formativo: 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:

metodologie di insegnamento: I docente introdurrà ogni argomento discutendo le questioni più rilevanti e le più interessanti e recenti evidenze sperimentali e applicazioni

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 funzionamento cognitivo del cervello umano (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 orale e progetto

Criteri di valutazione:

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

Testi di riferimento:

CONTENUTO NON PRESENTE

HUMAN COMPUTER INTERACTION

Titolare: Prof. LUCIANO GAMBERINI

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

Periodo: Il anno, 1 semestre

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

Tipologie didattiche: 42A; 6,00

LAW AND DATA

Titolare: Dott.ssa ELISA SPILLER

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Cognitive, Social and Economic 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

Curriculum: Piano di studio Machine Learning for Intelligence systems

BIG DATA COMPUTING

Titolare: Prof. ANDREA ALBERTO PIETRACAPRINA

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

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Machine Learning for Intelligence systems

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Il corso ha i seguenti prerequisiti: competenze relative al progetto e all'analisi di algoritmi e strutture dati, conoscenza delle nozioni fondamentali di calcolo delle probabilità e statistica, e capacità di programmazione in Java o Python.

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. Programming frameworks: MapReduce, Apache Spark Reducing input size (Case study: clustering) Reducing output size (Case study: frequent itemsets) Streaming framework.

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:

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. GIORGIO VALLE

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

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Machine Learning for Intelligence systems

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

BIOLOGICAL DATA

Titolare: Prof. SILVIO TOSATTO

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Machine Learning for Intelligence systems

Tipologie didattiche: 48A; 6,00

BUSINESS ECONOMIC AND FINANCIAL DATA

Titolare: Prof.ssa MARIANGELA GUIDOLIN

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Machine Learning for Intelligence systems

Tipologie didattiche: 48A; 6,00

COGNITION AND COMPUTATION

Titolare: da definire

Periodo: I anno, 1 semestre

Indirizzo formativo: Piano di studio Machine Learning for Intelligence 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 l'apprendimento cooperativo e le discussioni interattive su domande aperte. Questo promuoverà l'apprendimento e l'abilità di riflettere in modo critico sui concetti discussi.

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 inoltre preparare inoltre un elaborato scritto individuale che approfondisce un argomento assegnato durante il corso e che dovrà essere consegnato il giorno dell'esame scritto.

Criteria 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:

Russell, S. J., and Norvig, P. Artificial Intelligence: A Modern Approach. 2nd ed. Prentice Hall/P, Artificial Intelligence: A Modern Approach. 3rd Edition. : Prentice Hall/Pearson Education, 2010 Goodfellow, I., Bengio, Y., and Courville, A. Deep Learning., Deep Learning. : MIT Press, 2016 Koller, D., and Friedman, N. . ., Probabilistic Graphical Models: Principles and Techniques. : MIT Press, 2009

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 (soprattutto articoli scientifici) forniti dal docente sulla piattaforma di e-learning.

COGNITIVE, BEHAVIORAL AND SOCIAL DATA

Titolare: Prof. GIUSEPPE SARTORI

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Machine Learning for Intelligence systems

Tipologie didattiche: 48A; 6,00

FUNDAMENTALS OF INFORMATION SYSTEMS

Titolare: Prof. GIORGIO MARIA DI NUNZIO

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Machine Learning for Intelligence systems

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

Criteria 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: Piano di studio Machine Learning for Intelligence systems

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 (es., MatLab, Python, C, Java); conoscenze in Analisi Matematica e Algebra Lineare; ogni ulteriore conoscenza dei processi di rete in economia, biologia, telecomunicazioni, semantica, etc. può essere utile.

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. Conoscere i principali ambiti di applicazione, anche interdisciplinari, delle tecniche studiate 7. Essere in grado di sintetizzare l'analisi di una rete in un documento professionale 8. Essere in grado di implementare al computer algoritmi atti all'analisi delle reti

Attività di apprendimento previste e metodologie di insegnamento:

Il modulo prevede: - 16 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 computer per l'analisi di una rete; - 4 lezioni interattive finalizzate a progetti interdisciplinari con gli studenti del corso gemello di Social Network Analytics.

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; Evolution networks. 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 di comunità - Dendrogrammi; Metodo di Girvan Newman e betweenness; Ottimizzazione della misura di modularità (Louvain modularity); Clustering spettrale; Consensus clustering; Misura della similarità tra diverse soluzioni; Algoritmi per comunità sovrapposte. 6. Rappresentazione di una rete: funzionalità grafiche di Gephi e R/Python; algoritmi force-directed per la visualizzazione grafica di una rete. 7. Twitter Lab - Come estrarre una rete semantica da Twitter.

Modalità di esame:

La verifica delle conoscenze e delle abilità attese viene effettuata tramite 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 computer (in un qualsiasi linguaggio di programmazione noto allo studente) degli algoritmi richiesti per l'analisi; la stesura di una elaborato; 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.

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 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:

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

Eventuali indicazioni sui materiali di studio:

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

PROCESS MINING

Titolare: Prof. MASSIMILIANO DE LEONI

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Machine Learning for Intelligence systems

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Conoscenze di base di algoritmi, strutture dati e programmazione, come acquisiti nel corso "Fundamental of Information Systems"

Conoscenze e abilità da acquisire:

Lo scopo del corso è quello di introdurre e studiare metodi e concetti che riguardano la modellazione e l'analisi di processi di business. In dettaglio, il corso inizia con illustrare i principi di modellazione di processi come ci si attende che vengano eseguiti (i cosiddetti "TO-BE processes"), usando i più comuni linguaggi di modellazione (BPMN, Petri Nets e Declare), e con descrivere le tecniche di analisi della loro correttezza. Successivamente, il corso si focalizza sul Process Mining, che si pone l'obiettivo di utilizzare lo storico dei dati transazionali (i log degli eventi) per analizzare e migliorare i processi come sono eseguiti (noti come "AS-IS processes"), che possono differire da quelli attesi. Un' enfasi speciale verrà messa sugli algoritmi e sulla loro implementazione in piattaforme software open-source e commerciali. Sessioni pratiche complementano il corso affinché gli studenti possano toccare con mano i problemi di modellazione e mining di processi aziendali in contesti reali/realistici, utilizzando dati reali.

Attività di apprendimento previste e metodologie di insegnamento:

Il corso consiste di lezioni, esercitazioni e sessioni di laboratorio. Le sessioni di laboratorio creeranno l'opportunità di "sporcarsi le mani" su dati reali, così da comprendere la complessità di eseguire un progetto di Process Mining e di lavorare con dati di processo reali.

Contenuti:

Il corso affronterà i seguenti argomenti: 1. MODELLAZIONE ED ANALISI ATTRAVERSO IL LINGUAGGIO BPMN - Modellazione di processo base ed avanzata in BPMN - Analisi Qualitativa - Analisi Quantitativa 2. MODELLAZIONE ED ANALISI 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 Petri Nets: Principi e Verifica. 3. 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 Inductive 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 Sociali 4. TECNICHE PREDITTIVE E PROCESSI DI BUSINESS - Tecniche di Base di Predictive Process Monitoring - Tecniche Avanzate di Predictive Process Monitoring

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. : Springer, 2013 Wil M.P. van der Aalst and Christian Stahl, Modeling Business Processes: A Petri Net-Oriented Approach.. : Information Systems, 2011 Wil M.P. van der Aalst, Process Mining: Data Science in Action. Berlin: Springer-Verlag, 2016

Eventuali indicazioni sui materiali di studio:

Slides, Eserciziari, Esercitazioni di Laboratori Autoguidate e Articoli Scientifici.

Curriculum: Piano di studio Mathematics of Data Science**BIOLOGICAL DATA**

Titolare: Prof. SILVIO TOSATTO

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Mathematics of Data Science

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 quattro parti, che corrispondono ai diversi tipi di dati biologici: 1) Sequenze 1.1) DNA e proteine 1.2) Banche dati 1.3) Allineamenti 2) Strutture 2.1) Folding proteico 2.2) Banche dati 2.3) Predizione di struttura 3) Letteratura 3.1) Articoli scientifici 3.2) Banche dati 3.3) Text mining 3.4) Funzione 4) Reti di interazione 4.1) Regioni non-globulari 4.2) Interazioni biologiche 4.3) Banche dati 4.4) Proprieta' emergenti

Modalità di esame:

L'esame si compone di tre parti separate, che devono essere superate tutte: (i valori tra parentesi indicano i pesi per il voto complessivo) 1) Compitino di meta' semestre (ca. 17%) Domande a risposta aperta 2) Progetto (ca. 50%) Programmazione e analisi dataset biologico; relazione sul progetto 3) Esame finale (ca. 33%) Orale, con domande su progetto e corso

Criteri di valutazione:

Viene valutata: 1) la comprensione di concetti e gli algoritmi presentati a lezione 2) la capacità di applicare le nozioni fornite a lezione su problemi reali 3) la capacità critica di saper utilizzare i metodi nei modi più opportuni, scegliendo tra le alternative possibili 4) la capacità di sviluppare software riutilizzabile estendendo librerie esistenti 5) 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 (appena disponibili) e le registrazioni audio (podcast), le dispense e la letteratura usata per i progetti.

BUSINESS ECONOMIC AND FINANCIAL DATA

Titolare: Prof.ssa MARIANGELA GUIDOLIN

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Basic statistics: descriptive statistics and probability. Inferential statistics: estimation, confidence intervals and hypothesis testing.

Conoscenze e abilità da acquisire:

This 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:

Lectures and Laboratories. Working in groups.

Contenuti:

Decomposing and analysing economic time series: latent component approaches and ARMA modelling. Enhancing the analysis of economic and financial time series data: some case studies. Business and marketing data analyses: the joint use of cross-sectional and temporal dimension and the introduction of dynamic modelling.

Modalità di esame:

Homework and final presentation

Criteri di valutazione:

Students will be evaluated according to their level of knowledge of some tools and techniques to analyse economic, financial or business data and their ability to apply them to real cases.

Testi di riferimento:

Fitzmaurice G.M., Laird N.M., Ware J.H., Applied longitudinal analysis. : Wiley, 2011 Leeflang P.S.H, Wittnk D.R., Naert P.A., Building Models for Marketing Decisions. : Springer, 2000 Martin V., Hurn S., Harris D., Econometric Modelling with Time Series: Specification, Estimation and Testing. : Cambridge University Press, 2012 Woolbridge J.M., Econometric Modelling with Time Series. : The MIT press, 2010 Hyndman, R. J., Athanasopoulos, G., Forecasting: principles and practice.. : OTexts., 2018 Azzalini, A., Scarpa, B., Data analysis and data mining: An introduction.. USA: OUP, 2012

Eventuali indicazioni sui materiali di studio:

Notes prepared by the teaching staff will be uploaded throughout the course.

COGNITIVE, BEHAVIORAL AND SOCIAL DATA

Titolare: Prof. GIUSEPPE SARTORI

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

GAME THEORY

Titolare: ELVINA GINDULLINA

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

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

Prerequisiti:

Un corso anche basilare di teoria della probabilità.

Conoscenze e abilità da acquisire:

L'insegnamento prevede l'acquisizione delle seguenti conoscenze e abilità, suddivise in due insiemi. 1: parte base. 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. 2: parte applicativa. Sapere applicare i concetti della teoria dei giochi a scenari pratici, specialmente di tipo ICT; in questo contesto, e' di particolare interesse l'abilità di contestualizzare la teoria dei giochi come strumento di valutazione per l'efficacia della risoluzione tramite procedure multi-agente distributed.

Attività di apprendimento previste e metodologie di insegnamento:

Lezioni convenzionali con il supporto di slide. Prevista interazione su piattaforma moodle. Caricamento video via Kaltura.

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 Giochi cooperativi, il nucleo, il valore di Shapley Allocazione delle risorse Utilità, scelte e paradossi Giochi potenziali, coordinazione Algoritmi bio-inspired Giochi evolutivi Reti cognitive Selfish routing Sistemi multi-input con teoria dei giochi

Modalità di esame:

In qualunque caso l'esame comprende un test scritto obbligatorio a libro aperto, dove vengono sottoposti diversi problemi di game theory allo studente su argomenti toccati durante il corso. Per ogni esercizio, vengono poste più domande, tipicamente tre. Per frequentanti, l'esame può coinvolgere lo sviluppo

di un progetto in gruppi di 1-3 persone, su argomenti del corso applicati alle ICT. L'adesione a questa modalita' e l'argomento del progetto sono concordati con il docente durante il corso. Se il test scritto e' sufficiente, si puo' registrare il voto conseguito come voto finale dell'esame. Si puo' ulteriormente discutere il progetto sviluppato durante il corso con un esame orale, da svolgersi dopo l'esame scritto. Questi esami orali si svolgono nella stessa giornata di un esame scritto, ma non necessariamente bisogna presentarsi nella stessa giornata per l'esame scritto e la discussione orale del progetto. La discussione orale integra il voto dello scritto.

Criteri di valutazione:

Ogni domanda nei test scritti viene valutata fino a un massimo di 3 punti. La discussione del progetto viene valutata fino a 10 punti. Il voto finale e' la somma numerica dei punteggi individuali delle domande e della discussione del progetto (se presente), limitata a 30. Un punteggio di 30 e lode e' assegnato agli studenti il cui punteggio numerico e' superiore a 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 capacita' di problem solving - la capacita' di discussione e verifica ex-post della soluzione trovata. Nella valutazione del progetto (se presente) vengono tenuti in considerazione: - l'originalita' della proposta e la pertinenza sia con le tematiche del corso che con le metodologie ingegneristiche tipiche dell'ICT - la qualita' dell'esposizione orale - la capacita' di lavoro di gruppo e la presenza di singoli contributi attribuibili ai partecipanti al progetto - la capacita' 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 Noam Nisan, Tim Roughgarden, Eva Tardos, Vijay V. Vazirani (eds.), Algorithmic Game Theory. : Cambridge Univ. Press, 2007 A. MacKenzie, L. DaSilva, Game Theory for Wireless Engineers. : Morgan&Claypool, 2006 Roberto Lucchetti, A Primer in Game Theory. : Esculapio, 2011

Eventuali indicazioni sui materiali di studio:

Diversi libri forniscono una trattazione generale di teoria dei giochi. A mero titolo di suggerimento, si puo' usare il libro di Tadelis come riferimento in senso generale. Questa parte comunque dovrebbe essere integrata con materiale per le applicazioni. Il libro di MacKenzie e DaSilva e' un buon esempio, anche se non e' obbligatorio usare un libro per questo scopo (si puo' fare riferimento anche a materiale trovato in rete). In ogni caso, il docente fornira' agli studenti tutte le dispense delle lezioni e appunti aggiuntivi.

HIGH DIMENSIONAL PROBABILITY FOR DATA SCIENCE

Titolare: Prof. MARCO FORMENTIN

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Mathematics of Data Science

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. D

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 sullo 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:

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

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: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Mathematics of Data Science

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:

NON FREQUENTANTI: L'esame sarà orale con 3 domande sul libro, una delle quali proposta come esercizio (vedere il libro per trovare esempi). In caso di permanenza della situazione di emergenza l'orale sarà svolto on-line FREQUENTANTI: L'esame sarà basato su un lavoro personale di ricerca da svolgere durante il corso come lavoro per casa e su una breve prova orale di presentazione e discussione del medesimo. Un report sintetizzerà il lavoro svolto in modo simile a un paper scientifico o a un report professionale. La ricerca da sviluppare verrà riassunta in un report che includerà i seguenti punti: 1 - introduzione, contestualizzazione teorica, benchmarking 2 - processo di design/co-design 3 - sviluppo/modifica prototipo (per chi ha meno esperienza tecnica ci sono soluzioni) 4 - valutazione - analisi dei dati (es: UX, usability, presence, acceptance) 5 - risultati - discussione finale Qualsiasi tecnologia potrà essere sviluppata/adottata per la ricerca purché sia interattiva (es: web, robot, realtà virtuale/aumentata, dispositivi smart home, strumenti di lavoro, tools di Arduino, veicoli, strumenti musicali, apps. su cellulare/Tablet).

Criteri di valutazione:

NON PARTECIPANTI: Ogni domanda ha lo stesso valore, in altri termini ogni domanda ha un peso di 10 su 30. Il punteggio finale sarà la somma del punteggio di ogni domanda. PARTECIPARE La valutazione si baserà in gran parte sulla qualità del report e delle sue parti con un punteggio di 5 punti per ciascuna di esse, per un punteggio massimo di 25. I restanti 5 punti saranno basati sulla presentazione orale della ricerca.

Testi di riferimento:

Helen Sharp, Jenny Preece, Yvonne Rogers, Interaction Design: Beyond Human-computer Interaction. : Wiley, 2019

Eventuali indicazioni sui materiali di studio:

FREQUENTANTI INTERACTION DESIGN (5th 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 (5th ed.)- di Helen Sharp, Jenny Preece e Yvonne Rogers è l'unico materiale da utilizzare per questo corso se non lo si frequenta.

HUMAN DATA ANALYTICS

Titolare: Prof. MICHELE ROSSI

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

LAW AND DATA

Titolare: Dott.ssa ELISA SPILLER

Mutuato da: Laurea magistrale in Data Science (Ord. 2017)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

Prerequisiti:

No prerequisites

Conoscenze e abilità da acquisire:

Il corso renderà gli studenti in grado di comprendere gli aspetti giuridici e di anticipare le conseguenze legali delle attività basate sull'I.A. Gli studenti saranno chiamati a riflettere sulle questioni più contemporanee e controverse dell'I.A.

Attività di apprendimento previste e metodologie di insegnamento:

Lezioni Seminari Workshops Letture preassegnate.

Contenuti:

- the concept of data; personal, sensitive and economic data; big data - the concepts of identity and digital identity - property of data, choices in the management of data - supranational, international and national laws on data processing - civil and criminal protection of privacy - new contents and concepts of privacy: big data, cell phones; videos; wearable technologies, etc. - the right to be forgotten - social network, right to be forgotten, responsibility - provider's criminal responsibility - civil and criminal aspects of profiling activity - automatic data processing, human responsibilities - big data (collection, analysis, processing) and their influence on fundamental rights - the issue of genetic data - big data and economy - phishing - financial crimes and artificial intelligence

Modalità di esame:

Esame scritto

Criteri di valutazione:

Gli studenti verranno valutati in base al grado di conoscenze teoriche e pratiche nei campi oggetto dell'attività didattica e in base alla loro capacità di riflettere criticamente sugli aspetti legalmente più controversi della I.A.

Testi di riferimento:

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio:

Appunti dalle lezioni. Materiali utili possono essere reperiti a questo indirizzo web: <http://www.bigdatalaw-unipd.it/>

MATHEMATICAL CELL BIOLOGY

Titolare: Prof. MORTEN GRAM PEDERSEN

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

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

NETWORK SCIENCE

Titolare: Prof. TOMASO ERSEGHE

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

Periodo: Il anno, 1 semestre

Indirizzo formativo: Piano di studio Mathematics of Data Science

Tipologie didattiche: 48A; 6,00

STATISTICAL METHODS FOR HIGH DIMENSIONAL DATA

Titolare: Prof. BRUNO SCARPA

Periodo: Il anno, 1 semestre

Indirizzo formativo: 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:

Lezioni frontali. Esercitazioni in laboratorio

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, p1, ERGM) 2.5 Bayesian models (Stochastic block models, Latent space

models) 3. COMPUTATIONAL ISSUES 4. TOPICS IN STATISTICAL LEARNING AND DATA MINING METHODS On the basis of the interests of the class some of the following methods may be presented 4.1 STATISTICAL METHODS FOR TEXT MINING 4.2 CLUSTERING 4.3 GENERALIZAZIONS OF BOOSTING 4.4 OTHER TOPICS

Modalità di esame:

Prova pratica e prova orale

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