



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

SCUOLA DI SCIENZE

Bollettino Notiziario

Anno Accademico 2018/2019

**Laurea magistrale in Data Science (Ord.
2017)**

Curriculum: Corsi comuni

ALGORITHMIC METHODS AND MACHINE LEARNING

(Titolare: Prof. ALESSANDRO SPERDUTI)

Periodo: I anno, 2 semestre
Indirizzo formativo: Corsi comuni
Tipologie didattiche: 96A; 12,00 CFU

Prerequisiti :

The student should have basic knowledge of programming.

Conoscenze e abilità da acquisire :

This class teaches the basics in algorithmic methods and machine learning.

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. The other part will focus on machine learning techniques typically used in a data science scenario. First of all, general concepts concerning machine learning will be introduced. Then, the theory underpinning each presented technique will be explained and subsequently followed by practical demonstrations using Python and Scikit-Learn.

Attività di apprendimento previste e metodologie di insegnamento :

The course consists of lectures.

Contenuti :

The course will cover the topics listed below:

- 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. Recurrence relations. Case study: sorting.

Dynamic programming: coping with repeating subproblems. Memoization of recursive code. Case study: optimization algorithms on sequences.

Greedy paradigm: solving by successive choices. Applicability of the paradigm. Case study: data compression.

- Machine Learning

Introduction to Machine Learning: why machine learning is useful; when to use it.; where to use it; Machine Learning paradigms; basic ingredients of Machine Learning; complexity of the hypothesis space; complexity measures; examples of supervised learning algorithms.

Application Issues: A classification pipeline, representation and selection of categorical variables; model selection, evaluation measures.

in Depth (theory and practice using Python and Scikit-Learn): Support Vector Machines; Decision Trees and Random Forest; Neural Networks and Deep Learning; Manifold Learning; Kernel Density Estimation.

Modalità di esame :

Written exam and (individual) project. The project is due by the end of the course.

Criteri di valutazione :

The project work, and the written exam, will be evaluated on the basis of the following criteria: i) student's knowledge of the concepts, methods, and technologies; ii) ability of the student to master the implementation technology; iii) student's capacity for synthesis, clarity, and abstraction, as demonstrated by the written exam and project presentation. 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 :

Log of the lessons, didactic material (slides) and the detailed exam modalities will be available in the Website accessible from the MOODLE platform.

BIG DATA COMPUTING

(Titolare: Prof. ANDREA ALBERTO PIETRACAPRINA) - Mutuato da:

Periodo: I anno, 2 semestre
Indirizzo formativo: Corsi comuni
Tipologie didattiche: 48A; 6,00 CFU

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 e attività propedeutiche allo svolgimento degli homework.

Contenuti :

Il corso affronterà i seguenti argomenti:

Introduzione al fenomeno dei Big Data

Programming frameworks: MapReduce/Hadoop, Spark

Clustering

Association Analysis

Graph Analytics (metriche di centralità, scale-free/Power-law graphs, fenomeno dello small world, uncertain graphs)

Similarity and diversity search

Modalità di esame :

L'esame consiste in alcuni homework di programmazione, assegnati ogni 2-3 settimane e da svolgere in gruppi di 3-4 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 sito web accessibile anche da MOODLE:

<http://www.dei.unipd.it/~capri/BDC/>

BIOINFORMATICS

(Titolare: Prof. GIORGIO VALLE) - Mutuato da: Laurea magistrale in Informatica (Ord. 2014)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Corsi comuni

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

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.

BIOINFORMATICS AND COMPUTATIONAL BIOLOGY

(Titolare: Prof. SILVIO TOSATTO) - Mutuato da:

Periodo: I anno, 2 semestre
Indirizzo formativo: Corsi comuni
Tipologie didattiche: 48A; 6,00 CFU

Prerequisiti :

Conoscenze base di bioinformatica, p.es. metodi di allineamento e database.

Conoscenze e abilità da acquisire :

Il corso intende comunicare conoscenze per metodi bioinformatici di analisi delle proteine. Inoltre, intende indurre lo studente a poter svolgere autonomamente ricerche in silico con strumenti bioinformatici disponibili.

Attività di apprendimento previste e metodologie di insegnamento :

Il corso si compone di lezioni frontali, esercitazioni pratiche al computer e journal club. Le esercitazioni sono da svolgere secondo le istruzioni fornite e complementate dallo studio di un problema bioinformatico diverso per ogni gruppo. Il journal club si articola in presentazioni di articoli della letteratura recente.

Contenuti :

- 1) Relazione evolutiva struttura/funzione/interazioni delle proteine
- 2) Teorie di folding ed evoluzione delle proteine
- 3) Predizione di struttura 3D per omologia e metodi ab initio; L'esperimento CASP
- 4) Predizione di caratteristiche strutturali
- 5) Predizione di funzione delle proteine; L'esperimento CAFA
- 6) Interazioni tra proteine
- 7) Cenni di Network Biology
- 8) Correlazione genotipo-fenotipo e proteine; L'esperimento CAGI.

Modalità di esame :

L'esame si compone di quattro parti separate, che devono essere superate tutte: (i valori tra parentesi indicano i pesi per il voto complessivo)

- 1) Valutazione delle esercitazioni (25%)
- 2) Presentazione journal club (25%)
- 3) Stesura di una relazione finale su una proteina ignota (25%)
- 4) Esame orale con discussione sulla relazione finale e domande sul corso (25%)

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 espositiva e di discussione critica durante il journal club

Testi di riferimento :

J. Gu, P.E. Bourne (Ed.), *Structural Bioinformatics*. : Wiley, 2009

Eventuali indicazioni sui materiali di studio :

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

BIOLOGICAL DATA

(Titolare: Prof. SILVIO TOSATTO)

Periodo: Il anno, 1 semestre
Indirizzo formativo: Corsi comuni
Tipologie didattiche: 48A; 6,00 CFU

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 strutturale, 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) Reti di interazione
 - 3.1) Interazioni biologiche
 - 3.2) Banche dati
 - 3.3) Proprieta' emergenti
- 4) Letteratura
 - 4.1) Articoli scientifici
 - 4.2) Banche dati
 - 4.3) Text mining

Modalita' 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) Compitini inerenti alle esercitazioni (ca. 20%)
- 2) Progetto (ca. 50%)
- 3) Presentazione del progetto con valutazione critica (ca. 30%)

Criteri di valutazione :

Viene valutata:

- 1) la comprensione di concetti e gli algoritmi presentati a lezione
- 2) la capacit  di applicare le nozioni fornite a lezione su problemi reali
- 3) la capacit  critica di saper utilizzare i metodi nei modi pi  opportuni, scegliendo tra le alternative possibili
- 4) la capacit  di sviluppare software riutilizzabile estendendo librerie esistenti
- 5) la capacit  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: Dott. MAURO BERNARDI)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Corsi comuni

Tipologie didattiche: 48A; 6,00 CFU

Prerequisiti :

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

Conoscenze e abilita' 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.

Attivita' 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.

Modalita' 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. and J.H. Ware, *Applied Longitudinal Analysis*. : Wiley, 2011

Leeflang P.S.H., Wittink D.R., Wedel M. and P.A. Naert, *Building Models for Marketing Decisions*. : Springer, 2000

Martin V., Hurn S. and D. Harris, *Econometric Modelling with Time Series: Specification, Estimation and Testing*. : Cambridge University Press, 2012

Wooldridge J.M., *Econometric Analysis of Cross Section and Panel Data*. : The MIT Press, 2010

Eventuali indicazioni sui materiali di studio :

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

COGNITIVE SERVICES

(Titolare: Prof. ALESSANDRO SPERDUTI) - Mutuato da: Laurea magistrale in Informatica (Ord. 2014)

Periodo: I anno, 2 semestre
Indirizzo formativo: Corsi comuni
Tipologie didattiche: 32A+16L; 6,00 CFU

Prerequisiti :

Lo studente deve avere conoscenze di base di programmazione e algoritmi. È inoltre consigliabile conoscere i concetti di base in termini di probabilità e di analisi delle funzioni multivariate.

Conoscenze e abilità da acquisire :

Questo corso insegna i concetti, i metodi e le tecnologie alla base dei Servizi Cognitivi, vale a dire API, SDK e servizi, tipicamente disponibili nella nuvola (cloud), che aiutano gli sviluppatori software a creare applicazioni di intelligenza artificiale. Esempi di funzioni intelligenti che possono essere aggiunte ad un'applicazione tramite l'utilizzo di Servizi Cognitivi sono: il rilevamento delle emozioni da video; riconoscimento facciale, del contenuto visivo e vocale; comprensione linguistica e del parlato.

Il corso inoltre insegna le competenze e le abilità specifiche necessarie per applicare tali concetti alla progettazione e all'implementazione di applicazioni di intelligenza artificiale.

Gli studenti dovranno affrontare esercizi pratici in laboratorio informatico, in modo da provare l'applicazione delle conoscenze acquisite a piccoli esempi pratici.

Attività di apprendimento previste e metodologie di insegnamento :

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

Contenuti :

Il corso comprende gli argomenti elencati di seguito:

- Introduzione:

Dalla conoscenza umana ai servizi cognitivi intelligenti; Breve introduzione ai paradigmi di Intelligenza Artificiale e Apprendimento Automatico.

- Servizi cognitivi:

Concetti basilari; Lingua, Discorso e servizi di visione; Servizi e API principali (IBM Watson, Microsoft, Google Cloud); Tecnologie abilitanti.

- Problemi di apprendimento automatico e di applicazione:

Classificazione; Apprendimento delle rappresentazioni e selezione delle variabili categoriali; Processo di apprendimento e di valutazione; Misure di valutazione.

- Riconoscimento visivo:

"Insegnare ai computer a vedere": estrarre informazioni ricche da dati visivi; Sfide: perché la visione artificiale è difficile?; Progettare funzionalità visive efficaci; Apprendimento delle rappresentazioni nella visione artificiale; Comprensione delle immagini.

- Esercizi pratici:

Cosa c'è nella scatola? Come costruire una pipeline di riconoscimento visivo; Utilizzo di servizi cognitivi per il riconoscimento / comprensione delle immagini; Combinazione di diversi servizi in uno scenario multi-modale.

Modalità di esame :

Lo studente deve sviluppare, in accordo con il docente, un piccolo progetto applicativo. Inoltre, lo studente deve presentare una relazione scritta sul progetto svolto, in cui si discutono criticamente tutte le questioni trattate durante la sua realizzazione. Lo studente presenterà e discuterà il progetto e, se ritenuto necessario dal docente, affronterà un esame orale.

Criteri di valutazione :

Il lavoro di progetto e l'eventuale esame orale saranno valutati sulla base dei seguenti criteri: a) conoscenza da parte dello studente dei concetti, dei metodi e delle tecnologie alla base dei Servizi Cognitivi; b) capacità dello studente di padroneggiare la tecnologia di implementazione; c) capacità di sintesi, chiarezza e astrazione dello studente, come dimostrato dalla relazione scritta e dal progetto.

Testi di riferimento :

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio :

Le presentazioni mostrate durante le lezioni sono rese disponibili agli studenti come materiale di riferimento.

COGNITIVE, BEHAVIORAL AND SOCIAL DATA

(Titolare: Prof. GIUSEPPE SARTORI)

Periodo: I anno, 1 semestre
Indirizzo formativo: Corsi comuni
Tipologie didattiche: 48A; 6,00 CFU

Prerequisiti :

Nozioni di apprendimento automatico

Conoscenze e abilità da acquisire :

Alla fine del corso gli studenti saranno in grado di comprendere problemi complessi nelle scienze cognitive, sociali e comportamentali, di scegliere le metodologie più appropriate per estrarre informazioni dai dati, e di integrare le conoscenze di data science con aspetti riguardanti le scienze sociali, il cervello, la mente e il comportamento. Verranno inoltre acquisiti:

- I concetti di base di psicologia cognitiva, psicologia sociale e scienze del comportamento.
- Gli strumenti e le metodologie dell'analisi dei dati cognitivi, comportamentali e sociali.
- Abilità pratiche di analisi dei dati applicata a problemi cognitivi, comportamentali e sociali.

Attività di apprendimento previste e metodologie di insegnamento :

Il 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 scritto e orale

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

COMPUTER AD NETWORK SECURITY

(Titolare: Prof. MAURO CONTI) - Mutuato da: Laurea magistrale in Informatica (Ord. 2014)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Corsi comuni

Tipologie didattiche: 40A; 6,00 CFU

Prerequisiti :

Il corso non prevede propedeuticità.

Tuttavia, sono consigliate conoscenze di base di reti, crittografia, e sistemi distribuiti (tipicamente acquisite nei corsi di Laurea in Informatica).

Conoscenze e abilità da acquisire :

Acquisire concetti di base di sicurezza (e.g., Access Control, User Authentication, Malware, Attacchi DoS, Intrusion Detection/Prevention, Software and OS security, Trusted Computing) e conoscenze di sicurezza di sistema in ambiente Linux/Windows/Android, sicurezza di reti wireless/wired, web-application security.

Al termine del corso gli studenti saranno in grado di: progettare l'architettura di sistemi ed applicazioni sicure, e aggiornare autonomamente le proprie competenze nel settore, anche tramite risultati recenti della ricerca nell'area.

Attività di apprendimento previste e metodologie di insegnamento :

Lezioni frontali; discussione di articoli scientifici.

Contenuti :

1) COMPUTER SECURITY TECHNOLOGY AND PRINCIPLES: Cryptographic Tools, User Authentication, Access Control, Database Security, Malicious Software, Denial-of-Service Attacks, Intrusion Detection, Firewalls and Intrusion Prevention Systems.

2) SOFTWARE SECURITY AND TRUSTED SYSTEMS: Buffer Overflow, Software Security, Operating System Security, Trusted Computing and Multilevel Security.

3) MANAGEMENT ISSUES: IT Security Management and Risk Assessment, IT Security Controls, Plans, and Procedures, Physical and Infrastructure Security, Human Resources Security, Security Auditing, Legal and Ethical Aspects.

4) PART FOUR CRYPTOGRAPHIC ALGORITHMS: Symmetric Encryption and Message Confidentiality, Public-Key Cryptography and Message Authentication.

5) NETWORK SECURITY: Internet Security Protocols and Standards, Internet Authentication Applications, Wireless Network Security.

The second part of the course takes the form of seminars based on a selection of scientific papers (that either have had a strong impact on security today, or explore novel ideas that may be important in the future).

Modalita' di esame :

Scritta.

Criteri di valutazione :

Conoscenza dei concetti studiati nel corso.

Testi di riferimento :W. Stallings, L. Brown, *Computer Security: Principles and Practice 2/E.* : Prentice Hall,M. Bishop, *Introduction to Computer Security.* : Addison-Wesley Professional,**Eventuali indicazioni sui materiali di studio :**Libro (testo principale *Computer Security: Principles and Practice 2/E*) e articoli scientifici.

Il corso sarÃ tenuto in Inglese.

Il sito web del corso offrirÃ tutte le informazioni e materiale ulteriore:

<http://www.math.unipd.it/~conti/teaching.html>

FINALE EXAMINATION

(Titolare: da definire)

Periodo: Il anno, 1 semestre**Indirizzo formativo:** Corsi comuni**Tipologie didattiche:** ; 15,00 CFU**Contenuti :**

Elaborato scritto preparato sulla base dell'attivita di Stage

Testi di riferimento :

CONTENUTO NON PRESENTE

FUNDAMENTALS OF INFORMATION SYSTEMS

(Titolare: Dott. GABRIELE TOLOMEI)

Periodo: I anno, 1 semestre**Indirizzo formativo:** Corsi comuni**Tipologie didattiche:** 96A; 12,00 CFU**Prerequisiti :**

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

Conoscenze e abilita' 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 datacenters.

Attivita' 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 storing, and it covers the following topics:

Introduction to relational databases: data model; relational algebra; SQL; DBMS;

NoSQL technologies: characteristics of NoSQL databases; aggregate data models: key value stores, document databases, column family stores, graph databases, others; distribution models: sharding, replication (master-slave, peer-to-peer).

Streams of Data: architecture(s); data modeling; query processing and optimization.

- 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).

Modalita' 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.

GAME THEORY

(Titolare: Prof. LEONARDO BADIA) - Mutuato da:

Periodo: Il anno, 1 semestre
Indirizzo formativo: Corsi comuni
Tipologie didattiche: 48A; 6,00 CFU

Prerequisiti :

Un corso anche basilare di teoria della probabilità .

Conoscenze e abilita' da acquisire :

L'insegnamento prevede l'acquisizione delle seguenti conoscenze e abilita' , 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'abilita' di contestualizzare la teoria dei giochi come strumento di valutazione per l'efficacia della risoluzione tramite procedure multi-agente distributed.

A seconda del tipo di fruizione/frequenza/mutuazione del corso da parte dello studente, l'importanza di queste due insiemi viene considerata diversamente. In particolare, la parte applicativa e' rilevante per studenti di corsi di ambito ingegneristico frequentanti. Per gli studenti di corsi differenti, o per studenti non frequentanti, l'acquisizione di conoscenze e abilita' si concentra nell'insieme degli obiettivi di base.

Attivita' di apprendimento previste e metodologie di insegnamento :

Lezioni convenzionali con il supporto di slide.

Prevista interazione su piattaforma moodle.

Contenuti :

Concetti base di teoria dei giochi
Utilita', 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, complessita', 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
Utilita', scelte e paradossi

Giochi potenziali, coordinazione
Algoritmi bio-inspired
Giochi evolutivi
Reti cognitive
Selfish routing
Sistemi multi-input con teoria dei giochi

Modalita' di esame :

Per studenti di ingegneria frequentanti, l'esame coinvolge (a differenza degli altri) lo sviluppo di un progetto in gruppi di 1-3 persone, su argomenti del corso applicati alle ICT. L'argomento del progetto e' concordato con il docente durante il corso.

Per tutti gli studenti in qualunque caso l'esame comprende anche un esame scritto obbligatorio a libro aperto, dove vengono sottoposti quattro diversi problemi di game theory allo studente su argomenti toccati durante il corso. Per ogni esercizio, vengono poste tre domande.

Gli studenti di ingegneria frequentanti si limiteranno a risolvere 3 esercizi sui 4 proposti. Tutti gli altri (studenti non frequentanti, o studenti non di ingegneria) dovranno risolverli tutti e 4.

Se il test scritto e' sufficiente, gli studenti non frequentanti o gli studenti non di ingegneria potranno registrare il voto conseguito come voto finale dell'esame.

Gli studenti di ingegneria frequentanti invece discuteranno 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 lo studente deve presentarsi nella stessa giornata per l'esame scritto e la discussione orale del progetto. Per il superamento dell'esame e' necessaria una valutazione sufficiente di entrambe le parti, esame scritto e discussione orale.

Criteri di valutazione :

Ogni domanda nei test scritti viene valutata fino a un massimo di 3 punti.

Per gli studenti di ingegneria frequentanti, 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

A. MacKenzie, L. DaSilva, *Game Theory for Wireless Engineers.* : Morgan&Claypool, 2006

Noam Nisan, Tim Roughgarden, Eva Tardos, Vijay V. Vazirani (eds.), *Algorithmic Game Theory.* : Cambridge Univ. Press, 2007

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.

HUMAN COMPUTER INTERACTION

(Titolare: Prof. LUCIANO GAMBERINI) - Mutuato da:

Periodo: Il anno, 1 semestre

Indirizzo formativo: Corsi comuni

Tipologie didattiche: 42A; 6,00 CFU

Prerequisiti :

Non sono richiesti particolari prerequisiti.

Conoscenze e abilita' da acquisire :

Il corso offre la possibilita' 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'usabilita' 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 :

- Limiti umani e loro implicazioni per il design di tecnologie
- Modelli di interazione e stili utente
- Paradigmi e strategie per la costruzione di sistemi interattivi
- Progettare per l'usabilità : metodi e tecniche di interaction design
- Valutazione dell'esperienza utente: parte I, laboratorio
- Valutazione dell'esperienza utente: parte II, il mondo reale
- Valutazione dell'esperienza utente: parte III, tecniche avanzate (Eye tracking, misure psicofisiologiche, etc)
- Accessibilità e progettazione universale
- Social computing
- Argomenti speciali e studi di casi (usabilità dei siti web, applicazioni mobili, ambienti di vita e smart city, realtà aumentata e virtuale, interazione persona-robot, tecnologie persuasive per il cambiamento del comportamento, ergonomia dell'e-health)
- Interazione simbiotica ed altri argomenti recenti di ergonomia e HCI

Modalità di esame :

Esame in forma scritta con 5 domande chiuse e due aperte.

La prima domanda aperta valuta la capacità acquisita dallo studente nell'utilizzare metodi e tecniche appresi nel corso per risolvere casi concreti di design o valutazione ergonomica di un sistema interattivo o delle sue interfacce.

La seconda domanda aperta valuta la capacità di sintetizzare e commentare teorie e metodi appresi durante il corso o sul manuale.

Le domande chiuse sono a scelta multipla e hanno lo scopo di verificare le conoscenze teoriche e metodologiche acquisite.

Criteri di valutazione :

15/30 punti MAX per la prima domanda aperta (uso di metodi e tecniche HCI)

5/30 punti MAX per la seconda domanda aperta (sintesi e commento)

2/30 punti per ogni domanda chiusa corretta

Testi di riferimento :

DIX, FINLAY, ABOWD, BEALE, HUMAN-COMPUTER INTERACTION. New York: Pearson, 2012

Eventuali indicazioni sui materiali di studio :

Il libro " ora disponibile nella quarta edizione cartacea, ma la terza edizione on-line gratuita " quella suggerita per il corso.

PDF: https://www.researchgate.net/publication/224927543_Human-Computer_Interaction

Molteplici risorse interattive, slide, esempi, esercizi sono disponibili all'indirizzo <http://www.hcibook.com/e3/plain/>

HUMAN DATA ANALYTICS

(Titolare: Prof. MICHELE ROSSI)

Periodo: I anno, 2 semestre

Indirizzo formativo: Corsi comuni

Tipologie didattiche: 48A; 6,00 CFU

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.

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 architetturali e le relative prestazioni

- Attività di laboratorio (12 ore): sei lezioni verranno dedicate all'attività di laboratorio, nella quale le tecniche sviluppate durante le lezioni teoriche verranno implementate e caratterizzate sperimentalmente utilizzando il linguaggio Python.

Tutto il materiale didattico presentato / utilizzato (slides e software) 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 di feature, feature statistiche, feature spettrali, feature correlate al contesto 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 :

Watt, Jeremy; Borhani, Reza, *Machine learning refined: risorse elettroniche foundations, algorithms, and applications* Jeremy Watt, Reza Borhani, Aggelos Katsaggelos. New York: Cambridge University Press, 2016
Bishop, Christopher M., *Pattern recognition and machine learning* Christopher M. Bishop. New York: Springer, 0
Bengio, Yoshua; Courville, Aaron; Goodfellow, Ian, *Deep Learning*. Cambridge: MIT 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.

INTRODUCTION TO OMIC DISCIPLINES

(Titolare: Prof.ssa MARIA PENNUTO)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Corsi comuni

Tipologie didattiche: 48A; 6,00 CFU

Prerequisiti :

Conoscenze di base della teoria dell'evoluzione, genetica e biologia molecolare.

Conoscenze e abilità da acquisire :

Gli studenti apprenderanno come vengono generati i dati di "omics", da quali tipi di campioni biologici derivano (DNA, RNA, proteine), nozioni da integrare con la teoria dell'evoluzione di Darwin e le leggi della genetica di Mendel. Si introdurrà il concetto di mutazione, da distinguere dal concetto di Single Nucleotide Polymorphism (SNP). Agli studenti verranno forniti esempi di come si ottengono i dati di omics in lab. Impareranno come l'evoluzione e la pressione selettiva modificano la genetica degli esseri viventi e la trasmissione dell'informazione genetica nell'equilibrio dinamico che governa i processi biologici. Gli studenti apprenderanno il dogma della biologia, che vede la conservazione dell'informazione genetica nel DNA, molecola da cui origina l'RNA, da cui a sua volta originano le proteine, e come l'informazione genetica è conservata nel nostro genoma e trasmessa alla progenie. Inoltre, gli studenti impareranno come e perché l'espressione genica differisce da un tessuto ad un altro, facendo sì che lo stesso gene venga differenzialmente espresso in tessuti diversi.

Attività di apprendimento previste e metodologie di insegnamento :

L'insegnante farà utilizzo di diapositive (file ppt) che saranno disponibili agli studenti sulla piattaforma moodle.

Si effettueranno ripassi al fine di approfondire specifici argomenti. Si faranno verifiche ad intervalli al fine di valutare lo stato di apprendimento degli studenti.

Contenuti :

In questo corso gli studenti inizieranno dalla teoria dell'evoluzione fino ad arrivare al concetto di gene, mutazioni, genetica e leggi di Mendel. Gli studenti otterranno nozioni di biologia molecolare al fine di capire da quali fonti -campioni biologici e materiale biologico (DNA, RNA, proteina)- si ottengono i dati di omics. Gli argomenti specifici trattati nel corso saranno i seguenti:

- 1) Teoria dell'evoluzione
- 2) Leggi di Mendel: L'inizio dell'era della genetica
- 3) Mutazioni e polimorfismi
- 4) La cellula: Prokarioti vs eukarioti
- 5) Organelli subcellulari: Nucleus, cytosol, mitocondri, reticulum endoplasmaticum/complesso di Golgi, lisosomi
- 6) Dogma: DNA, RNA, proteina, dal gene alla proteina
- 7) Dati di OMICS dal DNA (genomics): Eterocromatina, eucromatina, DNA codificante vs non-codificante, replicazione del DNA
- 8) Dati di OMICS dall'RNA (transcriptomics): Trascrizione, splicing, microRNA, lncRNA
- 9) Dati di OMICS dalle proteine (Proteomics): Il codice genetico
- 10) Tecniche di biologia molecolare per processare il DNA: Sequenziamento di Sanger, metodi di sequenziamento di nuova generazione (NGS), PCR, clonaggio ed espressione genica
- 11) Tecniche di biochimica: Analisi delle proteine via spettrometria di massa, immunoprecipitazione, interazioni proteina-proteina, analisi di stabilit  delle proteine, analisi del metaboloma
- 12) Effetto della specie, tessuto, et  , sesso sui dati omics
- 13) Effetto dell'ambiente sui dati di OMICS
- 14) Uso dei dati OMICS come nuova chiave di lettura della teoria dell'evoluzione

Modalit  di esame :

Esame orale: Allo studente verr  chiesto di presentare un argomento a piacere. Seguiranno due domande specifiche. Lo studente potr  fare uso di diapositive sull'argomento a scelta.

Criteri di valutazione :

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

Testi di riferimento :

Lewin, Benjamin; Krebs, Jocelyn E.; Kilpatrick, Stephen T.; Goldstein, Elliott S., *Lewin's genes* 12.edited by Jocelyn E. Krebs, Elliott S. Goldstein, Stephen T. Kilpatrick. Burlington (MA): Jones & Bartlett Learning, 2017

Eventuali indicazioni sui materiali di studio :

Le diapositive utilizzate saranno fornite agli studenti.

KNOWLEDGE AND DATA MINING

(Titolare: Dott. LUCIANO SERAFINI)

Periodo: I anno, 2 semestre

Indirizzo formativo: Corsi comuni

Tipologie didattiche: 48A; 6,00 CFU

Prerequisiti :

Suggested basic knowledge of logics and statistics.

Conoscenze e abilita' da acquisire :

Introduce the students to the principles for logics for knowledge representation and reasoning, statistical relational learning, and the combination of the two in order to build system for learning and reasoning in hybrid domains.

Attivit  di apprendimento previste e metodologie di insegnamento :

Lectures supported by exercises and lab

Contenuti :

(A) Logics for knowledge representation:

(A.i) introduction to propositional logics, syntax, semantics, decision procedure. Satisfiability, weighted satisfiability, and best satisfiability.

(A.ii) First order logics, syntax, semantics, resolution and unification.

(A.iii) Fuzzy logics, syntax, semantics, and reasoning.

(B) statistical relational learning:

(B.i) Graphical models

(B.ii) Markov Logic Networks

(B.iii) Probabilistic prolog,

(B.iii) Logic Tensor Networks

Modalit  di esame :

Final examination based on: written examination or project development.

Criteri di valutazione :

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

Testi di riferimento :

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio :

Lecture notes and slides for the part not covered by textbooks will be provided.

LAW AND DATA

(Titolare: Dott.ssa SILVIA SIGNORATO)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Corsi comuni

Tipologie didattiche: 48A; 6,00 CFU

Prerequisiti :

No prerequisites

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 :

Oral Exam

Testi di riferimento :

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio :

Notes will be provided by the teacher

MATHEMATICAL MODELS AND NUMERICAL METHODS FOR BIG DATA

(Titolare: Dott. STEFANO CIPOLLA)

Periodo: Il anno, 1 semestre

Indirizzo formativo: Corsi comuni

Tipologie didattiche: 48A; 6,00 CFU

Prerequisiti :

Background on Matrix Theory: Type of matrices: Diagonal, Symmetric, Normal, Positive Definite; Matrix canonical forms: Diagonal, Schur; Matrix spectrum: Kernel, Range, Eigenvalues, Eigenvectors and Eigenspaces Matrix Factorizations: LU, Cholesky, QR, SVD

Conoscenze e abilità da acquisire :

Learning the mathematical and computational foundations of state-of-the-art numerical algorithms that arise in the analysis of big data and in many machine learning applications. By using modern Matlab toolboxes for large and sparse data, the students will be guided through the implementation of the methods on real-life problems arising in network analysis and machine learning.

Attività di apprendimento previste e metodologie di insegnamento :

Lectures supported by exercises and lab

Contenuti :

Numerical methods for large linear systems

— Jacobi and Gauss-Seidel methods — Subspace projection (Krylov) methods — Arnoldi method for linear systems (FOM) — (Optional) Sketches of GMRES — Preconditioning: Sparse and incomplete matrix factorizations

Numerical methods for large eigenvalue problems

— The power method — Subspace Iterations — Krylov-type methods: Arnoldi (and sketches of Lanczos + Non-Hermitian Lanczos) — (Optional) Sketches of their block implementation — Singular values VS Eigenvalues — Best rank-k approximation

Large scale numerical optimization

â—| Steepest descent and Newton's methods â—| Quasi Newton methods: BFGS â—| Stochastic steepest descent â—| Sketches of inexact Newton methods â—| Sketches Limited memory quasi Newton method

Network centrality

â—| Perron-Frobenius theorem â—| Centrality based on eigenvectors (HITS and Pagerank) â—| Centrality based on matrix functions

Data and network clustering

â—| K-Means algorithm â—| Principal component analysis and dimensionality reduction â—| Laplacian matrices, Cheeger constant, nodal domains â—| Spectral embedding â—| (Optional) Lovasz extension, exact relaxations, nonlinear power method (sketches)

Supervised learning

â—| Linear regression â—| Logistic regression â—| Multiclass classification â—| (Optional) Neural networks (sketches)

Modalità di esame :

Written exam

Testi di riferimento :

CONTENUTO NON PRESENTE

METHODS AND MODELS FOR COMBINATORIAL OPTIMIZATION

(Titolare: Dott. LUIGI DE GIOVANNI) - Mutuato da: Laurea magistrale in Informatica (Ord. 2014)

Periodo: Il anno, 1 semestre
Indirizzo formativo: Corsi comuni
Tipologie didattiche: 32A+4E+12L; 6,00 CFU

Prerequisiti :

Elementi di ricerca operativa, elementi di programmazione lineare, elementi di base di programmazione.

Conoscenze e abilità da acquisire :

Uso di metodologie avanzate di supporto alle decisioni per la modellazione e la soluzione di problemi di ottimizzazione combinatoria. Il corso intende fornire strumenti matematici e algoritmici per la soluzione di problemi pratici di ottimizzazione con l'utilizzo dei pacchetti software e delle librerie di ottimizzazione più diffuse.

Attività di apprendimento previste e metodologie di insegnamento :

L'insegnamento prevede lezioni frontali, esercitazioni in laboratorio e discussione di esempi notevoli. Le esercitazioni in laboratorio consistono nell'implementazione di algoritmi di ottimizzazione combinatoria sia esatti (con l'uso di librerie di programmazione lineare intera) sia euristici).

Contenuti :

1. Approfondimenti e applicazioni di Programmazione Lineare e dualità : metodo del simplesso primale-duale, tecniche di generazione di colonne, applicazioni a problemi di ottimizzazione su grafo.
2. Metodi avanzati di Programmazione Lineare Intera (PLI): Branch & Bound e tecniche di rilassamento, formulazioni alternative di modelli PLI, metodo dei piani di taglio e tecniche di Branch & Cut, applicazioni ad esempi notevoli: commesso viaggiatore, problemi di localizzazione, problemi di network design etc.
3. Meta-euristiche di Ottimizzazione Combinatoria: ricerca di vicini e varianti, algoritmi evolutivi.
4. Applicazione di metodi di modellazione e ottimizzazione su grafo.
5. Laboratori: utilizzo di software e librerie di ottimizzazione.

Modalità di esame :

Esame orale sui contenuti del corso. Realizzazione facoltativa di un progetto individuale sulla soluzione di un problema, reale o realistico, di ottimizzazione combinatoria (definizione del problema, modellazione, applicazione di un metodo di soluzione esatto e/o euristico).

Criteri di valutazione :

L'esame verifica il livello di apprendimento degli argomenti svolti e la capacità dello studente di applicarli per la soluzione di problemi reali di ottimizzazione combinatoria.

Testi di riferimento :

CONTENUTO NON PRESENTE

Eventuali indicazioni sui materiali di studio :

Dispense fornite dal docente. Articoli scientifici.

OPTIMIZATION FOR DATA SCIENCE

(Titolare: Prof. FRANCESCO RINALDI)

Periodo: I anno, 2 semestre
Indirizzo formativo: Corsi comuni
Tipologie didattiche: 48A; 6,00 CFU

Prerequisiti :

Conoscenze di base in

- Analisi reale e Calcolo;
- Algebra lineare;
- Teoria della probabilità.

Conoscenze e abilità da acquisire :

Comprendere modelli e metodi di ottimizzazione nell'ambito del Data Science. Nello specifico:

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

Attività di apprendimento previste e metodologie di insegnamento :

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

Contenuti :

1. Ottimizzazione lineare: Teoria e algoritmi

- (a) Modelli di programmazione lineare in Data Science;
- (b) Dualità;
- (c) Metodo del simplesso;
- (d) Metodo dei punti interni;

2. Insiemi convessi e funzioni convesse

- (a) Convessità, nozioni di base;
- (b) Funzioni convesse: nozioni di base e proprietà;

3. Ottimizzazione convessa non vincolata:

- (a) Modelli in Data Science;
- (b) Caratterizzazione delle soluzioni ottime;
- (c) Metodi tipo gradiente;
- (d) Metodi tipo gradiente a blocchi;
- (e) Metodi per l'ottimizzazione stocastica.

4. Ottimizzazione convessa vincolata;

- (a) Modelli in Data Science;
- (b) Caratterizzazione delle soluzioni ottime;
- (c) Metodi basati su approssimazioni poliedrali;
- (d) Metodi di proiezione;

5. Ottimizzazione su reti di grande dimensione

- (a) Modelli di reti in Data Science;
- (b) Metodi di ottimizzazione distribuita.

Modalità di esame :

- Prova scritta alla fine del corso
- Esercizi
- Progetto (Opzionale)

1) Gli Esercizi verranno assegnati periodicamente. Gli studenti dovranno consegnare le soluzioni a scadenze prestabilite.

2) La prova d'esame scritta prevede 5 domande a risposta aperta.

3) Il progetto (opzionale) può essere richiesto per approfondire tematiche specifiche.

La prova scritta contribuisce all' 85% del voto.

Gli esercizi contribuiscono al 15% del voto.

Il progetto permette un incremento da 1 a 3 punti del voto.

Criteri di valutazione :

La valutazione della preparazione dello studente si baserà :

- sulla comprensione degli argomenti svolti in aula;
- sull'acquisizione dei concetti di carattere teorico;
- sulla capacità di utilizzare in maniera autonoma e consapevole i modelli e le metodologie risolutive proposte.

Testi di riferimento :

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

Eventuali indicazioni sui materiali di studio :

- Dispense e slide fornite dal docente.

PROCESS MINING

(Titolare: Dott.ssa CHIARA GHIDINI)

Periodo: Il anno, 1 semestre
Indirizzo formativo: Corsi comuni
Tipologie didattiche: 48A; 6,00 CFU

Prerequisiti :

Basic notions of algorithms, data structures and programming.

Conoscenze e abilita' da acquisire :

The aim of the course is to introduce and investigate the main methods and concepts that pertain the modeling and analysis of business processes. More in detail, the course will focus on the main modeling languages (BPMN, Petri Nets, and Declare), on the main methodologies for manual modeling and analysis and on the main algorithms for the (semi)automatic modeling and analysis (the so-called process mining). By exploiting concrete software platforms several algorithms will be also investigated in a 'hands-on' fashion on real data.

At the end of the course the students should have a detailed knowledge of the main methods and concepts of business process modeling e mining, of the main metrics used to support the analysis of business processes and of the main algorithms of process mining.

Attività di apprendimento previste e metodologie di insegnamento :

The course consists of lectures. Some practical activities and exercises will require the use of computers.

Contenuti :

The course will cover the topics listed below:

1. MODELING AND ANALISYS: THE BPMN PERSPECTIVE

- Process Identification
- Essential and Advanced Process Modeling in BPMN
- Qualitative Analysis
- Quantitative Analysis
- Process redesign

2. MODELING AND ANALISYS: THE PETRI NET PERSPECTIVE

- An introduction to Petri Nets
- Petri nets and colored petri nets
- Simulation based analysis
- Reachability and coverability analysis
- Process modeling and analysis with PN

3. PROCESS MINING

- Data & Process mining
- Getting the data: the construction of event logs
- An introduction to Process discovery
- Advanced process discovery
- Conformance checking - replay based
- Conformance checking - logic based
- Mining additional perspectives
- Typical use cases, e.g., medical processes

4. DECLARATIVE APPROACHES

- Declarative approaches and Declare
- Declarative process mining (discovery in Declare) and hybrid approaches

5. PREDICTIVE PROCESS MONITORING

- Basic Predictive Process Monitoring techniques
- Advanced Predictive Process Monitoring techniques

Modalita' di esame :

Written exam and project. The project is due and has to be discussed by the end of the course.

Criteri di valutazione :

The project work, and the written exam, will be evaluated on the basis of the following criteria: i) student's knowledge of the concepts, methods, and technologies; ii) ability of the student to master the implementation technology; iii) student's capacity for synthesis, clarity, and abstraction, as demonstrated by the written exam and project presentation. The final grade is obtained as the weighted sum of the grades of the written exam (80%) and the project (20%)

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
W. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Berlin: Springer-Verlag, 2011

Eventuali indicazioni sui materiali di studio :

Slides, exercises and scientific papers will be provided.

STAGE

(Titolare: Prof. PAOLO DAI PRA)

Periodo: Il anno, 1 semestre
Indirizzo formativo: Corsi comuni
Tipologie didattiche: ; 15,00 CFU

Contenuti :

Lo stage consiste in un periodo di attivit  all'interno di un'azienda, un'istituzione o un centro di ricerca esterno all'Universit  di Padova

Testi di riferimento :

CONTENUTO NON PRESENTE

STATISTICAL LEARNING

(Titolare: da definire)

Indirizzo formativo: Corsi comuni

Moduli del C.I.:

Statistical Learning 1
Statistical Learning 2

STATISTICAL LEARNING 1

(Titolare: Prof. ALBERTO ROVERATO)

Periodo: I anno, 1 semestre
Indirizzo formativo: Corsi comuni
Tipologie didattiche: 48A; 6,00 CFU

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

Attivit  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

STATISTICAL LEARNING 2

(Titolare: Prof. ALBERTO ROVERATO)

Periodo: I anno, 2 semestre
Indirizzo formativo: Corsi comuni
Tipologie didattiche: 48A; 6,00 CFU

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

Attivit  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, Trevor J.; Tibshirani, Robert, *The elements of statistical learning data mining, inference, and prediction* Trevor Hastie, Robert Tibshirani, Jerome Friedman. New York: Springer, 2009
Lavine, M., *Introduction to Statistical Thought.* : None, 2013

STOCHASTIC METHODS

(Titolare: Prof. PAOLO DAI PRA)

Periodo: I anno, 1 semestre
Indirizzo formativo: Corsi comuni
Tipologie didattiche: 48A; 6,00 CFU

Prerequisiti :

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

Conoscenze e abilità da acquisire :

Lo scopo del corso è di introdurre metodi e concetti di Calcolo delle Probabilità che hanno un forte impatto come strumenti algoritmici, computazionali e nello studio delle reti. Attraverso l'uso del software R (R development Core Team, 2006), alcuni problemi specifici saranno trattati con simulazione al calcolatore.

Attività di apprendimento previste e metodologie di insegnamento :

Lezioni frontali. Alcune esercitazioni prevedono simulazioni al calcolatore

Contenuti :

1. Richiami di Calcolo delle Probabilità .

• distribuzioni discrete e continue

• variabili aleatorie, valor atteso e valor atteso condizionale

• approssimazione di distribuzioni di probabilità .

2. Catene di Markov e passeggiate aleatorie

• Catene di Markov e relative distribuzioni stazionarie

• Monte Carlo (MCMC), convergenza di algoritmi MCMC-based

• Electrical networks.

3. Grafi aleatori

• Grafi di Erdos-Renyi: connettività, componente gigante

• Grafi aleatori regolari

• Grafi dinamici. Preferential attachment.

Modalità di esame :

Esame scritto

Criteri di valutazione :

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

Testi di riferimento :

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

Eventuali indicazioni sui materiali di studio :

Il docente fornirà delle dispense, assieme ad altro materiale didattico: esercizi, eventuali estratti di articoli scientifici

STRUCTURAL BIOINFORMATICS

(Titolare: Dott. DAMIANO PIOVESAN) - Mutuato da: Laurea magistrale in Informatica (Ord. 2014)

Periodo: I anno, 2 semestre
Indirizzo formativo: Corsi comuni
Tipologie didattiche: 32A+16E; 6,00 CFU

Prerequisiti :

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

Conoscenze e abilità da acquisire :

Il corso intende comunicare le conoscenze di base sulla struttura e funzione della materia vivente nonché i principali metodi computazionali per il loro studio. Inoltre intende permettere allo studente lo svolgimento autonomo di un progetto di ricerca in bioinformatica strutturale, 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 del software implementato.

Contenuti :

Il corso si compone di due parti:

1) Introduzione alla materia vivente (2 CFU):

1.1) Cenni di chimica organica, interazioni deboli ed energetica

1.2) Struttura e funzione di DNA e proteine

1.3) Lipidi, membrane e trasporto cellulare

1.4) Metodi sperimentali per la determinazione strutturale

2) *Biochimica computazionale (4 CFU):*

2.1) *Banche dati biologiche*

2.2) *Librerie software e concetti per allineamenti di sequenza e ricerca in banche dati*

2.3) *Relazione sequenza e struttura nelle proteine e classificazione strutturale*

2.4) *Metodi per la predizione della struttura delle proteine da sequenza, lâ€™esperimento CASP*

2.5) *Metodi per la predizione di funzione ed interazioni delle proteine, lâ€™esperimento CAFA*

2.6) *Proteine non globulari, disordine e ripetizioni strutturali*

Modalita' 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) *Test scritto sulle nozioni di biochimica (ca. 30%)*

2) *Progetto software (ca. 40%)*

3) *Presentazione del progetto con valutazione critica (ca. 30%)*

Criteri di valutazione :

Viene valutata:

1) *la comprensione di concetti e gli algoritmi presentati a lezione*

2) *la capacitÃ di applicare le nozioni fornite a lezione su problemi reali*

3) *la capacitÃ critica di saper utilizzare i metodi nei modi piÃ opportuni, scegliendo tra le alternative possibili*

4) *la capacitÃ di sviluppare software riutilizzabile estendendo librerie esistenti*

5) *la capacitÃ espositiva e di discussione critica*

Testi di riferimento :

K.C. Mathews, K.E. Van Holde, K.G. Ahern, Biochimica (3Ã° edizione). : Casa Editrice Ambrosiana, 2004

Pascarella, Stefano; Paiardini, Alessandro, Bioinformaticadalla sequenza alla struttura delle proteine. Bologna: Zanichelli, 2011

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), le dispense e la letteratura usata per i progetti. Le dispense scaricabili in formato PDF contengono oltre 300 pagine per facilitare lo studio.