



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA

**SCUOLA DI SCIENZE**

**Bollettino Notiziario**

Anno Accademico 2017/2018

**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 :**

Competenze relative al progetto e all'analisi di algoritmi e strutture dati, e conoscenza delle nozioni fondamentali di calcolo delle probabilità e statistica.

**Conoscenze e abilità da acquisire :**

Il corso fornisce allo studente conoscenze relative ai principali strumenti e metodi per l'analisi di insiemi di dati potenzialmente grandi.

### **Attività di apprendimento previste e metodologie di insegnamento :**

Lezioni frontali e attività propedeutiche allo svolgimento del progetto

### **Contenuti :**

Il corso affronterà i seguenti argomenti:

Introduzione al fenomeno dei Big Data

Programming frameworks: MapReduce/Hadoop, Spark

Association Analysis

Clustering

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

Similarity and diversity search

### **Modalità di esame :**

Prova scritta e progetto (di gruppo) obbligatori. I progetti sono presentati e discussi a fine corso o, facoltativamente, in una prova orale successiva al superamento dello scritto.

### **Criteri di valutazione :**

La valutazione finale "A" ottenuta combinando la valutazione del progetto e quella dello scritto.

### **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 saranno resi disponibili sul sito web accessibile da 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 lezioni 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.

---

## **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:** l'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

â€¢ Misure comportamentali e come ottenerle (es. RT); misure di comportamento implicite ed esplicite (es. la IAT)

â€¢ Estrarre e predire informazioni dal comportamento (es. lie detection, predizione di "malicious behavior" dall'attività sui social networks, fake online reviews, security, ecc.)

â€¢ Misure psicofisiologiche e come ottenerle (es. HR variability, SCR, espressioni facciali, EEG, fMRI, etc.)

â€¢ Estrarre e predire informazioni dalle misure psicofisiologiche

â€¢ 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.)

â€¢ 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.)

â€¢ 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

## 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 programming and algorithms.

**Conoscenze e abilità da acquisire :**

This class teaches the concepts, methods, and technologies at the basis of storage, networking, and processing of data and big data. 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 part concerning networking 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. The programming part focuses on programming for data scientists using Python, starting from the description of its interactive computational environment, and continuing with storage, data manipulation, and visualization.

**Attività di apprendimento previste e metodologie di insegnamento :**

The course consists of lectures.

**Contenuti :**

The course will cover the topics listed below:

- Databases

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

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).

- Programming

Programming for Data Scientist using Python: computational environment (IPython and Jupyter); storage and manipulation (NumPy and Pandas); data visualization (Matplotlib).

**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.

# HUMAN DATA ANALYTICS

(Titolare: Prof. MICHELE ROSSI)

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

## Prerequisiti :

Si consiglia un corso di base che fornisca le basi della teoria delle probabilità. Un corso di programmazione di base è altresì utile per il corretto apprendimento dei concetti esposti.

## Conoscenze e abilità da acquisire :

Le competenze che verranno acquisite dallo studente sono molteplici e tutte collegate all'analisi di segnali biometrici o comunque generati dall'attività dell'uomo quali il suo movimento. In particolare, lo studente apprenderà le seguenti conoscenze tecniche:

- Il trattamento di segnali quasi-periodici, estraendone strutture ricorrenti tramite rimozione del rumore e segmentazione: alcuni segnali fondamentali quali l'elettrocardiogramma e segnali di movimento misurati da dispositivi indossabili (wearables) verranno analizzati in dettaglio;
- L'estrazione di "features" da insiemi di dati biometrici;
- L'applicazione di algoritmi di classificazione (clustering) a segnali biometrici al fine di classificare i soggetti che li hanno generati o costruire dizionari per una rappresentazione accurata ma al tempo stesso compatta dei dati stessi;
- L'utilizzo di tecniche di apprendimento non-supervisionato, basate su reti neurali, al fine di partizionare i dati in cluster caratteristici ("vector quantization") e costruire dizionari per una loro rappresentazione efficiente;
- L'utilizzo di costrutti statistici, quali le catene di Markov nascoste e le reti Bayesiane, per la modellizzazione di dati correlati;
- L'utilizzo di reti neurali come strumenti per la mappatura di patterns e per la soluzione di problemi di classificazione.

## Attività di apprendimento previste e metodologie di insegnamento :

L'apprendimento avverrà tramite lezioni frontali svolte dal titolare del corso. Il progetto finale servirà per mettere in pratica le tecniche viste a lezione e verificarne l'efficacia in presenza di segnali reali.

## Contenuti :

Parte I "Introduzione al corso"

- Intro: course outline, grading rules, office hours, etc.
- Applicazioni: salute, servizi "activity-aware", applicazioni di sicurezza e gestione delle emergenze, sistemi di autenticazione, analisi di dinamiche di attività umane

Parte II - Strumenti e tecniche

- Vector quantization (VQ):
  - Obiettivi, metriche di qualità
  - Algoritmi di VQ non-supervisionati:
    - Self-Organizing Maps (SOM), Time Adaptive-SOM (TASOM)
    - Gas Neural Networks (GNG)
- Reti Neurali "Deep" (DNN)
  - Neural networks brief: concept, examples, training
  - Reti Neurali Convolutionali: architettura, allenamento
- Modelli per l'analisi sequenziale dei dati:
  - Modelli di Markov nascosti (Hidden Markov Models, HMM):
    - Massima verosimiglianza per HMM
    - Forward-backward algorithm
    - Algoritmo somma-prodotto
    - Algoritmo di Viterbi
  - Teoria della decisione sequenziale di Wald's (caso a processi iid)

Parte III "Applicazioni (usando gli strumenti della Parte II)"

- Apprendimento di attività umana
  - Activities & sensors: definizioni, classi di attività
  - Features: features temporali, features statistiche, features spettrali, features estratte dal contesto
  - Riconoscimento di attività : segmentazione, finestre mobili, segmentazione non-supervisionata, metriche prestazionali e risultati
- Processamento di dati biometrici:
  - Segnali fisiologici quasi-periodici
    - Apprendimento di dizionari e algoritmi di compressione del segnale
  - Segnali inerziali: riconoscimento dell'identità
  - Sensori inerziali e video camere: tracciamento della camminata e sua profilazione
- Riconoscimento vocale:
  - Sistemi ibridi formati da reti neurali "deep" e modelli di Markov nascosti
  - Architettura, allenamento delle reti, prestazioni

## Modalità di esame :

L'esame consisterà nella presentazione di un elaborato (progetto) nel quale verrà risolto un problema di classificazione su un determinato dataset. Lo studente dovrà implementare il software per la classificazione in oggetto, ottenere i risultati e commentarli.

## Criteri di valutazione :

I seguenti criteri di valutazione verranno presi in considerazione:

- Qualità della presentazione orale;

- Qualit  del progetto finale: chiarezza espositiva, rigore matematico;
- Qualit  dei risultati ottenuti;
- Originalit  dell'approccio scelto;
- Complessit  del problema affrontato.

**Testi di riferimento :**

Bishop, Christopher M., *Pattern recognition and machine learning* Christopher M. Bishop. New York: Springer, 0  
 Cook, Diane J.; Krishnan, Narayanan C., *Activity learning* discovering, recognizing and predicting human behavior from sensor data Diane J. Cook, Narayanan C. Krishnan. Hoboken: NJ, Wiley, 2015  
 Jeremy Watt, Reza Borhani, Aggelos K. Katsaggelos, *Machine Learning Refined: Foundations, Algorithms, and Applications.* : Cambridge University Press, 2016  
 Goodfellow, Ian; Courville, Aaron, *Deep learning* Ian Goodfellow, Yoshua Bengio and Aaron Courville. Cambridge: MA [etc.], MIT Press, 2016

**Eventuali indicazioni sui materiali di studio :**

Il Professore fornir  tutto il materiale necessario che consister  primariamente in:

- 1) capitoli di libro (diversi libri verranno utilizzati, in dipendenza dell'argomento trattato);
- 2) articoli scientifici;
- 3) slides.

Tutto il materiale scritto sar  in lingua Inglese. Il materiale ai punti 1), 2) e 3) sar  reso disponibile tramite il sito del corso (protetto da password).

## 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 :**

, . . .

**Eventuali indicazioni sui materiali di studio :**

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

## 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.

**Conoscenze e abilita' da acquisire :**

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

**Attivit  di apprendimento previste e metodologie di insegnamento :**

Il corso si baser  su lezioni frontali.

**Contenuti :**

1. Ottimizzazione lineare: Teoria e algoritmi

(a) Modelli di programmazione lineare in Data Science;

(b) Dualit ;

(c) Metodo del Simplex;

(d) Metodo dei punti interni;

2. Insiemi convessi e funzioni convesse

(a) Convessit , nozioni di base;

(b) Funzioni convesse: nozioni di base e propriet ;

3. Ottimizzazione convessa non vincolata:

(a) Modelli in Data Science;

(b) Caratterizzazione delle soluzioni ottime;

(c) Metodi tipo gradiente;

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 clustering.

**Modalit  di esame :**

- Prova scritta alla fine del corso

- Progetto (Opzionale)

**Criteri di valutazione :**

La valutazione della preparazione dello studente si baser :

- sulla comprensione degli argomenti svolti in aula;

- sull'acquisizione dei concetti di carattere teorico;

- sulla capacit  di utilizzare in maniera autonoma e consapevole i modelli e le metodologie risolutive proposte.

**Testi di riferimento :**

CONTENUTO NON PRESENTE

**Eventuali indicazioni sui materiali di studio :**

- Dispense fornite dal docente.

- I testi di consultazione verranno indicati dal docente durante il corso

---

**STATISTICAL LEARNING**

(Titolare: da definire)

**Indirizzo formativo:** Corsi comuni

**Prerequisiti :**

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

**Conoscenze e abilit  da acquisire :**

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

**Modalit  di esame :**

written test

**Criteri di valutazione :**

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

**Moduli del C.I.:**

Statistical Learning 1

Statistical Learning 2

---

**STATISTICAL LEARNING 1**

(Titolare: Prof.ssa MONICA CHIOGNA)

**Periodo:** l'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

### **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

## **STATISTICAL LEARNING 2**

---

(Titolare: Prof.ssa MONICA CHIOGNA)

**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

### **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, 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  
â€¢ Catene di Markov e relative distribuzioni stazionarie  
â€¢ Monte Carlo (MCMC), convergenza di algoritmi MCMC-based  
â€¢ Electrical networks.

3. Grafi aleatori  
â€¢ Grafi di Erdos-Renyi: connettivit  , componente gigante  
â€¢ Grafi aleatori regolari  
â€¢ Grafi dinamici. Preferential attachment.

**Modalita' di esame :**

Esame scritto

**Criteri di valutazione :**

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

**Testi di riferimento :**

P. Dai Pra, *Stochastic Methods for Data Science.* , 2017

**Eventuali indicazioni sui materiali di studio :**

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

## STRUCTURAL BIOINFORMATICS

(Titolare: Prof. SILVIO TOSATTO) - 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. Linguaggi di programmazione C++ e/o Java.

**Conoscenze e abilita' da acquisire :**

Il corso intende comunicare le conoscenze di base sulla struttura e funzione della materia vivente nonch  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.

**Attivit  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
- 1.2) Interazioni deboli ed energetica
- 1.3) Struttura e funzione di DNA e proteine
- 1.4) Lipidi, membrane e trasporto cellulare

2) Biochimica computazionale (4 CFU):

- 2.1) Banche dati biologiche
- 2.2) Librerie software e concetti per allineamenti di sequenza, profili e ricerca in banche dati
- 2.3) Relazione sequenza   struttura   funzione nelle proteine e classificazione
- 2.4) Metodi per la predizione della struttura delle proteine da sequenza. L'esperienza CASP.
- 2.5) Metodi per la predizione di funzione delle proteine. L'esperienza CAFA.
- 2.6) Cenni di biologia delle reti e dei sistemi.
- 2.7) Correlazione genotipo   fenotipo. L'esperienza CAGI.

**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 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 :**

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

S. Pascarella, A. Paiardini, *Bioinformatica.* : 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) e le

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