



# UNIVERSITÀ DI PAVIA

Anno Accademico 2021/2022

## APPRENDIMENTO COMPUTAZIONALE IN BIOMEDICINA

<b>Anno immatricolazione</b>	2021/2022
<b>Anno offerta</b>	2021/2022
<b>Normativa</b>	DM270
<b>SSD</b>	ING-INF/06 (BIOINGEGNERIA ELETTRONICA E INFORMATICA)
<b>Dipartimento</b>	DIPARTIMENTO DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE
<b>Corso di studio</b>	BIOINGEGNERIA
<b>Curriculum</b>	Sanita' digitale
<b>Anno di corso</b>	1°
<b>Periodo didattico</b>	Secondo Semestre (07/03/2022 - 17/06/2022)
<b>Crediti</b>	9
<b>Ore</b>	80 ore di attività frontale
<b>Lingua insegnamento</b>	Italiano
<b>Tipo esame</b>	SCRITTO
<b>Docente</b>	BELLAZZI RICCARDO - 6 CFU RAMAT STEFANO - 3 CFU
<b>Prerequisiti</b>	Vengono richieste conoscenze di base sulla teoria delle probabilità e di statistica. Per la parte pratica, viene richiesta una certa dimestichezza con l'uso degli strumenti informatici.
<b>Obiettivi formativi</b>	Il corso si propone di fornire allo studente competenze metodologiche e tecniche per: - impiegare in ambito biomedico una vasta classe di algoritmi che sono in grado di apprendere regole decisionali da dati e di migliorare automaticamente le loro prestazioni sulla base dell'esperienza. Nella prima parte del corso verranno introdotte le problematiche generali del machine learning e i metodi più importanti nelle applicazioni. Lo studente, alla fine della prima parte del corso, lo studente dovrà essere in grado di: - impiegare un percorso metodologicamente adeguato per apprendere modelli decisionali dai dati - utilizzare strumenti software e pacchetti statistici di apprendimento

automatico. Il corso conterrà sia lezioni teoriche che esercitazioni pratiche al computer.

Nella seconda parte gli studenti verranno poi introdotti a due metodologie ampiamente utilizzate nell'ambito dell'Intelligenza Artificiale: reti neurali e deep learning da un lato e gli algoritmi genetici dall'altro.

Le prime rappresentano uno strumento di apprendimento computazionale per compiti di riconoscimento e classificazione sia statici che dinamici, i secondi sono un metodo di ottimizzazione a base stocastica di estrema versatilità. Per questa parte, al termine del corso gli studenti dovranno essere in grado di implementare in ambiente Matlab le principali architetture di reti "shallow" e "deep" per la classificazione e l'approssimazione, nonché algoritmi genetici generazionali e steady state.

#### Programma e contenuti

Prima parte. Apprendimento di regole decisionali

Introduzione al corso: Apprendimento automatico e Data mining nelle scienze biomediche.

Ambiti di applicazione dei metodi automatici per la classificazione: diagnosi, prognosi, ricerca

I concetti di base: esempi, istanze, attributi e rappresentazione delle regole decisionali

Apprendimento supervisionato

Alberi decisionali: apprendimento, tecniche di pruning

Metodi bayesiani: Naive Bayes, analisi discriminante

Modelli di regressione: modello lineare, regressione logistica, reti neurali, support-vector machines

Metodo k-nearest e misure di distanza

Random forest, boosting

Tecniche di feature selection. Information gain e Relief

Valutazione degli algoritmi di apprendimento e problemi di valutazione in ambito biomedico

Training e Testing. Accuratezza, calibrazione, specificità e sensibilità, precisione e recall, misura F

Metodi per la valutazione delle prestazioni. Cross validazione, Bootstrap e Curve ROC.

Apprendimento non supervisionato

Regole di associazione

Metodi di clustering: K-means, K-medoids, clustering gerarchico, Self-organizing maps

Valutazione dei risultati dei metodi di clustering

Metodologie per il data mining in bio-medicina

Applicazioni del data mining in bio-medicina: diagnosi, prognosi, classificazione, genomica funzionale

La metodologia CRISP per il data mining in bio-medicina.

Esercitazioni e laboratori

Uso dei software Orange, Weka e Matlab per la soluzione di problemi di classificazione.

Soluzione di problemi su data set forniti durante il corso.

Seconda parte.

Introduzione alle reti neurali.

Il perceptrone e adaline, reti basate su un singolo neurone per classificazione e approssimazione lineare.

Perceptrone multistrato e radial basis function networks.

	<p>Self organizing maps per clustering non supervisionato.</p> <p>Reti dinamiche: la rete di Hopfield, la rete di Elman e le sue evoluzioni, la state-space model network. Reti ricorrenti e Long Short Term Memory network. Reti convoluzionali, autoencoder, reti generative.</p> <p>Algoritmi genetici per l'ottimizzazione guidata da una funzione di utilità (fitness).</p>
<b>Metodi didattici</b>	<p>Lezioni di teoria, esercitazioni in classe e esercitazioni al computer.</p> <p>Disponibilità di video lezioni.</p>
<b>Testi di riferimento</b>	<p>T. Mitchell. Machine Learning. Mc Graw Hill..</p> <p>P. Tan, M. Steinbach, V. Kumar. Introduction to data mining. Addison Wesley..</p> <p>I. Witten, E. Frank. Data mining. Morgan Kaufmann.</p> <p>Riccardo Bellazzi. Slides delle lezioni.</p> <p>S. Haykin, Neural Networks and Learning Machines, Prentice Hall, 3rd Ed., 2009</p> <p>D.E. Goldberg, Genetic Algorithms in search, optimization and machine learning, Addison Wesley, 1989</p> <p>Stefano Ramat. Slides delle lezioni.</p>
<b>Modalità verifica apprendimento</b>	<p>Esame scritto e discussione di due relazioni (apprendimento e reti neurali) con analisi dati</p>
<b>Altre informazioni</b>	
<b>Obiettivi Agenda 2030 per lo sviluppo sostenibile</b>	<p><a href="#">\$lbl_legenda_sviluppo_sostenibile</a></p>