UNIVERSITÀ DI PAVIA

Anno Accademico 2020/2021

APPRENDIMENTO COMPUTAZIONALE IN BIOMEDICINA

Anno immatricolazione	2020/2021
Anno offerta	2020/2021
Normativa	DM270
SSD	ING-INF/06 (BIOINGEGNERIA ELETTRONICA E INFORMATICA)
Dipartimento	DIPARTIMENTO DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE
Corso di studio	BIOINGEGNERIA
Curriculum	Sanita' digitale
Anno di corso	1°
Periodo didattico	Secondo Semestre (08/03/2021 - 14/06/2021)
Crediti	9
Ore	80 ore di attività frontale
Lingua insegnamento	Italiano
Tipo esame	SCRITTO
Docente	BELLAZZI RICCARDO - 6 CFU RAMAT STEFANO - 3 CFU
Prerequisiti	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.
Obiettivi formativi	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 calcolatore.
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 sensitività, 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 percettrone e adaline, reti basate su un singolo neurone per classificazione e approssimazione lineare.
Percettrone multistrato e radial basis function networks.

	Self organizing maps per clustering non supervisionato. 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. Algoritimi genetici per l'ottimizzazione guidata da una funzione di utilità (fitness).
Metodi didattici	Lezioni di teoria, esercitazioni in classe e esercitazioni al computer. Disponibilità di video lezioni.
Testi di riferimento	T. Mitchell. Machine Learning. Mc Graw Hill.. P. Tan, M. Steinbach, V. Kumar. Introduction to data mining. Addison Wesley.. I. Witten, E. Frank. Data mining. Morgan Kaufmann. Riccardo Bellazzi. Slides delle lezioni. S. Haykin, Neural Networks and Learning Machines, Prentice Hall, 3rd Ed., 2009 D.E. Goldberg, Genetic Algorithms in search, optimization and machine learning, Addison Wesley, 1989 Stefano Ramat. Slides delle lezioni.
Modalità verifica apprendimento	Esame scritto e discussione di due relazioni (apprendimento e reti neurali) con analisi dati
Altre informazioni	
Obiettivi Agenda 2030 per lo sviluppo sostenibile	$lbl_legenda_sviluppo_sostenibile