

DENOMINAZIONE INSEGNAMENTO / ATTIVITÀ FORMATIVA

BIG DATA, MACHINE LEARNING AND ASTROPHYSICAL DATA

INFORMAZIONI INSEGNAMENTO / ATTIVITÀ FORMATIVA

A.A.	2026-2027	CdS	FISICA (LM-17)
Codice	8068923-8067554	Canale	Unico
CFU	4	Lingua	Inglese

DOCENTE RESPONSABILE DELL'INSEGNAMENTO / ATTIVITÀ FORMATIVA

Hervé Bourdin

Modulo:

OBIETTIVI FORMATIVI E RISULTATI DI APPRENDIMENTO ATTESI

Italiano	OBIETTIVI FORMATIVI: Nata come scienza osservativa, l'astrofisica si è spesso nutrita dal dialogo fruttuoso tra modellistica empirica e fisicamente motivata delle leggi della natura. Dovendo gestire insiemi di dati sempre più ricchi e voluminosi, gli astronomi devono oggi combinare diversi meccanismi fisici complessi e multi-scala tra di loro per interpretare le loro osservazioni. L'apprendimento automatico le aiuta ad esplorare le relazioni tra gli osservabili, la strutturazione dei dati oppure la trama dei campi stocastici. Il corso introduce alcuni concetti di statistica di grandi dimensioni, teoria dell'informazione e apprendimento automatico insieme al loro contesto matematico. Nei laboratori computazionali, si illustrano alcune applicazioni per osservazioni astronomiche. CONOSCENZA E CAPACITÀ DI COMPRENSIONE: Il corso comprende lezioni e laboratori riguardo alla riduzione della dimensionalità, il clustering, la teoria dei grafi, i Gaussian mixtures, la classificazione, la regressione (regolarizzata) e il model fitting, i processi Gaussiani, le reti neurali e la wavelet scattering transform. CAPACITÀ DI APPLICARE CONOSCENZA E COMPRENSIONE: Approfondita comprensione delle tecniche di acquisizione dei dati ed estrazione di informazioni. - Buona conoscenza dello stato dell'arte della strumentazione presentemente in uso in ambito astrofisico. AUTONOMIA DI GIUDIZIO:
-----------------	---

	Capacità di eseguire ricerche bibliografiche in tecniche astrofisiche e strumentazione, selezionando i materiali interessanti e valutandone i risultati principali. Capacità di discernere quale sia il miglior metodo per la misura di una sorgente astrofisica. ABILITÀ COMUNICATIVE: Capacità di presentare ed organizzare l'esposizione di un argomento specialistico di approfondimento di tecniche astrofisiche e strumentazione. CAPACITÀ DI APPRENDIMENTO: Capacità di affrontare nuovi campi attraverso uno studio autonomo.
English	LEARNING OUTCOMES: <i>As an observational science, astrophysics often benefited from a constructive dialog between empirical and physically motivated modellings of the laws of nature. Facing data sets of increasing size and richness, modern astronomers routinely have to combine various complex multi-scale physical mechanisms with each other to interpret their observations. Machine learning helps us explore the relationship between observables, the structuration of data sets or the texture of stochastic fields. The course introduces concepts of high dimension statistics, information theory and machine learning with their mathematical context. Computational laboratories illustrate applications to astronomical observations.</i> KNOWLEDGE AND UNDERSTANDING: <i>The course includes lectures and tutorials about data dimensionality reduction, clustering, graph theory, Gaussian mixtures, classification, (regularized) regression and model fitting, Gaussian processes, (convolutional) neural networks and scattering wavelet transforms.</i> APPLYING KNOWLEDGE AND UNDERSTANDING: <i>Understanding the limits of the instrumentation. - Ability to unassistedly analyze astrophysical data sets.</i> MAKING JUDGEMENTS: <i>Ability to perform bibliographic research on astrophysical techniques and instrumentation, selecting interesting materials and evaluating the main results. Ability to understand which is the best approach to perform an astrophysical measure.</i> COMMUNICATION SKILLS: <i>Ability to present and organize the exposition of a specialized topic of astrophysical methods and instrumentation.</i>

	LEARNING SKILLS: <i>To be able to approach new fields through independent study.</i>
--	---

PREREQUISITI

Italiano	Conoscenze di base di fisica generale e di analisi matematica.
English	<i>Basic knowledge of general physics and mathematical analysis.</i>

PROGRAMMA E CRONOPROGRAMMA

Italiano	• <i>Introduzione -- Cenni di teoria delle probabilità e variabile casuale: correlazione e dipendenza, distribuzioni di Gauss e del χ^2, convergenza stocastica</i> • <i>Clustering gerarchico -- Gaussian mixtures, teoria dei grafi</i> • <i>Struttura dei dati e riduzione della dimensionalità -- statistica in grandi dimensioni, connessione con la dipendenza probabilistica, componenti principali e processi stazionari</i> • <i>Regressione e model fitting empirico: soluzioni regolarizzate, accuratezza del fit e criteri informativi, regressione tramite processi Gaussiani</i> • <i>Classificazione generativa e discriminativa, alberi decisionali -- cenni di teoria dell'informazione, entropia di Shannon, divergenza di Kullback-Leibler</i> • <i>Apprendimento profondo: reti neurali artificiali, scattering wavelet transforms</i>
English	• <i>Introduction -- Notions of probability theory and random variables: correlation and dependence, Gauss and χ^2 distributions, stochastic convergence</i> • <i>Hierarchical clustering -- Gaussian mixtures, graph theory</i> • <i>Data structure and dimensionality reduction -- large dimension statistics, connection with probabilistic dependence, principal components and stationary processes</i> • <i>Regression and (empirical) model fitting: regularized solutions, fitting accuracy and information criteria, Gaussian process regression</i> • <i>Generative and discriminative classification, decision trees -- notions of information theory, Shannon entropy, Kullback-Leibler divergence</i> • <i>Deep learning: artificial neural networks, scattering wavelet transforms</i>

TESTI ADOTTATI E BIBLIOGRAFIA DI RIFERIMENTO

Italiano	Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data, Updated Edition Željko Ivezić, Andrew J. Connolly, Jacob T. VanderPlas, and Alexander Gray
English	Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for

	the Analysis of Survey Data, Updated Edition Željko Ivezić, Andrew J. Connolly, Jacob T. VanderPlas, and Alexander Gray
--	--

DESCRIZIONE DELLE MODALITÀ DI SVOLGIMENTO E METODI DIDATTICI ADOTTATI

☒ In presenza ☐ A distanza

Italiano	Lezioni frontali in aula, uso della lavagna e altro materiale didattico
<i>English</i>	<i>Classroom lessons, chalkboard and other didactic material</i>

MODALITÀ DI FREQUENZA

☐ frequenza obbligatoria ☒ frequenza facoltativa

DESCRIZIONE DELLE MODALITÀ DI FREQUENZA

Italiano	Lezioni in aula
<i>English</i>	<i>Class lectures</i>

MODALITÀ DI VALUTAZIONE

☐ Prova scritta ☒ Prova orale ☐ Prova di laboratorio ☐ Prova pratica
☐ Valutazione in itinere ☒ Valutazione di progetto ☐ Valutazione di tirocinio

DESCRIZIONE DELLE MODALITÀ E DEI CRITERI DI VERIFICA DELL'APPRENDIMENTO

Italiano	L'esame consiste in un colloquio orale volto a verificare il livello di conoscenza degli argomenti trattati. Per superare l'esame occorre riportare un voto non inferiore a 18/30; più in particolare, la prova di esame sarà valutata secondo i seguenti criteri: Non idoneo: importanti carenze e/o inaccuratezza nella conoscenza e comprensione degli argomenti; limitate capacità di analisi e sintesi, frequenti generalizzazioni. 18-20: conoscenza e comprensione degli argomenti appena sufficiente con possibili imperfezioni; capacità di analisi, sintesi e autonomia di giudizio sufficienti. 21-23: Conoscenza e comprensione degli argomenti routinaria; Capacità di analisi e sintesi corrette con argomentazione logica coerente. 24-26: Discreta conoscenza e comprensione degli argomenti; buone capacità di analisi e sintesi con argomentazioni espresse in modo rigoroso. 27-29: Conoscenza e comprensione degli argomenti completa; notevoli capacità di analisi, sintesi. Buona autonomia di giudizio.
-----------------	---

	30-30L: Ottimo livello di conoscenza e comprensione degli argomenti. Notevoli capacità di analisi e di sintesi e di autonomia di giudizio. Argomentazioni espresse in modo originale.
English	<i>The exam consists of an oral interview aimed at verifying the level of knowledge of the topics covered in class.</i> <i>To pass the exam, a grade of not less than 18/30 must be obtained; more specifically, the vote is determined as follows:</i> <i>18-21, if the student has acquired the basic concepts of the discipline, with sufficiently correct and appropriate language;</i> <i>22-25, if the student has acquired the basic concepts of the discipline in depth, and is adequately able both to make connections between the various topics and to explain concepts with appropriate and correct language;</i> <i>26-29, if the student has a complete and well-organized background of knowledge, being able both to apply and re-elaborate independently, without any error, the acquired knowledge, and to highlight logical-analytical skills with an appropriate and rigorous mathematical language;</i> <i>30 and 30 cum laude, if the student has a complete and in-depth knowledge that can be applied to complex cases and new situations as well; the student must explain the topics brilliantly with perfect language properties, as well as mastering the mathematical tools at the basis of the discipline.</i>