

Syllabus Attività Formativa

Anno Offerta	2022
Corso di Studio	B33 - DATA SCIENCE
Regolamento Didattico	B33-22-22
Percorso di Studio	GEN - GENERICO
Insegnamento/Modulo	A423694 - NUMERICAL METHODS FOR DATA SCIENCE -
Attività Formativa Integrata	-
Partizione Studenti	-
Periodo Didattico	S2 - Secondo Semestre
Sede	
Anno Corso	1
Settore	MAT/08 - ANALISI NUMERICA
Tipo attività Formativa	B - Caratterizzante
Ambito	70296 - Formazione matematico-statistica
CFU	9.0
Ore Attività Frontali	72.0
AF_ID	178659

Tipo Testo	Codice Tipo Teste	Num. Max. Caratteri	Ob bl.	Testo in Italiano	Testo in Inglese
-------------------	--------------------------	----------------------------	---------------	--------------------------	-------------------------

Lingua insegnamento	LINGUA_INS	100	Sì	Inglese	English
Contenuti	CONTENUTI	2000	Sì	<p>Il corso affronta i metodi numerici necessari per affrontare diverse classi di problemi nella moderna analisi dei dati e nell'apprendimento automatico come:</p> <ul style="list-style-type: none"> - Classificazione (classificatore PCA, LDA, LVQ,SVM) - analisi dei dati mediante grafi - analisi di regressione <p>Applicazione dei metodi a problemi di estrazione del testo, estrazione di immagini e page ranking.</p>	<p>The course covers numerical methods required for dealing with several classes of problems in modern data analysis and machine learning such as:</p> <ul style="list-style-type: none"> - Classification (PCA, LDA, LVQ, SVM classifiers) - Graph data analysis - Regression Analysis (Lasso, Model parameter identification) <p>Application to text mining, image mining, and page ranking</p>
Testi di riferimento	TESTI_RIF		Sì	<ol style="list-style-type: none"> 1. "Data Mining: An Algorithmic Approach to Clustering and Classification", by D. Calvetti and E. Somersalo (draft version) 2. Appunti del corso 	<ol style="list-style-type: none"> 1. "Data Mining: An Algorithmic Approach to Clustering and Classification", by D. Calvetti and E. Somersalo (draft version) 2. Lecture notes

Obiettivi formativi	OBIETT_FORM	3000	Sì	<p>Conoscenze: gli studenti devono acquisire una solida conoscenza di metodologie e conoscenze su metodi numerici e algoritmi per l'analisi dei dati.</p> <p>Applicazioni delle conoscenze e competenze: gli studenti dovrebbero essere in grado di selezionare e applicare correttamente i metodi numerici e gli strumenti software per l'analisi dei dati.</p> <p>Abilità comunicative: gli studenti dovrebbero essere in grado di illustrare i metodi e gli strumenti appresi durante il corso e di comunicare i risultati ottenuti, utilizzando un linguaggio tecnico e scientifico adeguato.</p>	<p>Knowledge and understanding: students are expected to acquire knowledge of numerical methods and algorithms for data analysis.</p> <p>Applying knowledge and understanding: students should be able to select and properly apply numerical methods and software tools for data analysis.</p> <p>Communication skills: students should be able to illustrate the methods and tools learned during the course and to communicate the results obtained with them, using a suitable technical and scientific language.</p>
Prerequisiti	PREREQ	2000	Sì	L'insegnamento non prevede propedeuticità, ma presuppone la conoscenza degli argomenti generalmente trattati in un corso di laurea triennale, tra i quali gli argomenti di base di ottimizzazione, algebra lineare e statistica.	the knowledge of foundation of optimization, numerical linear algebra and statistical computing

Metodi didattici	METODI_DID	2000	Sì	<p>Il corso prevede lezioni frontali (48 ore – 6 CFU) e sessioni di laboratorio (36 ore – 3 CFU) La frequenza ai corsi non è obbligatoria, ma è fortemente consigliata.</p>	<p>The course consists of lectures (48 hours, 6 CFU - ECTS credits) and laboratory sessions (36 hours, 3 CFU - ECTS credits).</p> <p>Course attendance is not mandatory, but it is strongly recommended.</p>
Altre informazioni	ALTRO	2000	Sì	<p>Le attività di laboratorio sono parte integrante del programma.</p>	<p>The laboratory activities are an integral part of the program.</p>
Modalità di verifica dell'apprendimento	MOD_VER_AP PR	3000	Sì	<p>Gli studenti vengono valutati attraverso una prova orale, volta a verificare il raggiungimento degli obiettivi del corso. Durante l'esame gli studenti devono presentare lo sviluppo di codici che implementino i metodi studiati.</p> <p>Per ciascuno degli algoritmi implementati, è necessaria l'esecuzione su una serie di problemi test, che evidenzino gli aspetti implementativi, le prestazioni dei codici implementati e l'analisi dei risultati ottenuti.</p> <p>A tale scopo, gli studenti possono utilizzare i programmi sviluppati da loro stessi o resi disponibili dal docente durante il corso.</p>	<p>Students are evaluated through an oral assessment, aimed at verifying if they matched the objectives of the course. During the assessment, students are also asked to provide a computer-based illustration of methods and tools studied in the course, through the execution is required on a set of test problems, which highlight the implementation aspects and the performance of the implemented codes, and the analysis of the results obtained.</p> <p>To this aim, students can use computer programs developed by themselves or made available by the teacher during the course.</p> <p>The use of other course material is not allowed.</p>

				<p>L'uso di altro materiale didattico non è consentito.</p> <p>I voti sono espressi in trentesimi. Il punteggio minimo richiesto è 18/30. Il voto massimo è 30/30 e Lode</p>	<p>Marks are expressed in the thirtieths. The minimum passing mark is 18/30. Outstanding performance is marked 30/30 cum laude.</p>
Programma esteso	PROGR_EST		Sì	<p>Argomenti trattati:</p> <p>Algoritmi di Classificazione:</p> <p>Concetti di base, classificatori, misure di similarità, I principali classificatori (PCA; LDA; LVQ; SVM)</p> <p>Analisi dei dati mediante grafi:</p> <p>Algoritmi di esplorazione mediante grafi; modelli random; modelli di Poisson</p> <p>Regressione:</p> <p>Regolarizzazione; norma l1; regolarizzazione di tipo Lasso</p> <p>Applicazione dei metodi a problemi di estrazione del testo, di immagini e pake ranking</p>	<p>Classification Algorithms:</p> <p>Basic concepts; Distance classifier; Alternative dissimilarity measures; k-nearest neighbor classifier; Principal Component Analysis (PCA) classifier; Linear Discriminant Analysis (LDA) classifier; Learning Vector Quantifier (LVQ); Support Vector Machine (SVM).</p> <p>Graph data analysis:</p> <p>Basic algorithms for graph exploration; Random graph models; Poisson random graphs; Generating functions; Poisson random graphs.</p> <p>Regression analysis:</p> <p>Regularization; l1 norm; lasso problem.</p> <p>Application of the previous methods to text mining, image mining, and page ranking.</p>

			<p>Attività di Laboratorio:Tutte le attività di laboratorio saranno sviluppate usando l'ambiente MATLAB</p> <p>Il programma dettagliato degli argomenti e delle attività di laboratorio sarà disponibile a fine corso</p>	<p>Laboratory activities: All the activities will be performed using the MATLAB software environment.</p> <p>The detailed program will be available at the end of the course</p>
--	--	--	---	--