

## Syllabus Attività Formativa

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

<b>Tipo Testo</b>	<b>Codice Tipo Teste</b>	<b>Num. Max. Caratteri</b>	<b>Ob bl.</b>	<b>Testo in Italiano</b>	<b>Testo in Inglese</b>

<b>Lingua insegnamento</b>	LINGUA_INS	100	Sì	Inglese	English
<b>Contenuti</b>	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"> <li>- Classificazione (classificatore PCA, LDA, LVQ,SVM)</li> <li>- analisi dei dati mediante grafi</li> <li>- analisi di regressione</li> </ul> <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"> <li>- Classification ( PCA, LDA, LVQ, SVM classifiers)</li> <li>- Graph data analysis</li> <li>- Regression Analysis (Lasso, Model parameter identification)</li> </ul> <p>Application to text mining, image mining, and page ranking</p>
<b>Testi di riferimento</b>	TESTI_RIF		Sì	<ol style="list-style-type: none"> <li>1. "Data Mining: An Algorithmic Approach to Clustering and Classification", by D. Calvetti and E. Somersalo (draft version)</li> <li>2. Appunti del corso</li> </ol>	<ol style="list-style-type: none"> <li>1. "Data Mining: An Algorithmic Approach to Clustering and Classification", by D. Calvetti and E. Somersalo (draft version)</li> <li>2. Lecture notes</li> </ol>

<b>Obiettivi formativi</b>	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>
<b>Prerequisiti</b>	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

<b>Metodi didattici</b>	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>
<b>Altre informazioni</b>	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>
<b>Modalità di verifica dell'apprendimento</b>	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>
<b>Programma esteso</b>	PROGR_EST		Sì	<p><b>Argomenti trattati:</b></p> <p><b>Algoritmi di Classificazione:</b></p> <p>Concetti di base, classificatori, misure di similarità, I principali classificatori (PCA; LDA; LVQ; SVM)</p> <p><b>Analisi dei dati mediante grafi:</b></p> <p>Algoritmi di esplorazione mediante grafi; modelli random; modelli di Poisson</p> <p><b>Regressione:</b></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><b>Classification Algorithms:</b></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><b>Graph data analysis:</b></p> <p>Basic algorithms for graph exploration; Random graph models; Poisson random graphs; Generating functions; Poisson random graphs.</p> <p><b>Regression analysis:</b></p> <p>Regularization; l1 norm; lasso problem.</p> <p>Application of the previous methods to text mining, image mining, and page ranking.</p>

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