Academic subject: Numerical Methods in Data Science

Degree Class: LM-35 (Mathematics)		**Degree Course:** Mathematics	**Academic Year:** 2020/2021
		Kind of class: optional	**Year:** third / **Period:** second semester
			ECTS: 7 divided into **ECTS lessons: 6.5** **ECTS exe/lab/tutor: 0.5**

Time management, hours, in–class study hours, out–of–class study hours

lesson: 52 exe/lab/tutor: 8 in–class study: 60 out–of–class study:113

Language: Italian	**Compulsory Attendance:** no		
Subject Teacher: Nicoletta Del Buono	**Tel:** +39 0805442711 **e–mail:** nicoletta.delbuono@uniba.it	**Office:** Department of Mathematics Room 24, Floor Second	**Office days and hours:** Monday 10:00-11:00 a.m. Other days: on appointment

Prerequisites:

General knowledges related to first degree in Mathematics (L-35 Mathematics) with particular emphasis to Numerical Methods and Multivariable Mathematical Analysis

Educational objectives:

Acquisition of basic numerical techniques for optimization of nonlinear multivariate functions and for solving linear programming problems. Acquisition of basic knowledge for exploratory Data Analysis and for using optimization mechanisms to deal with problems arising in learning from data

Expected learning outcomes (according to Dublin Descriptors)

Knowledge and understanding:

Acquiring some techniques for solving classification and clustering problems.
Ability to using numerical codes implementing standard techniques for analyzing real data.

Applying knowledge and understanding:

Theoretical and practical knowledge acquired would be used on applied mathematical fields and for solving problems in the context of learning from data.

Making judgements:

Ability to identify the right numerical techniques which is able to address and solve numerically optimization problems arising in real-life applications using large structured datasets

Communication:

Acquiring the advanced mathematical language and formalism required for the consultation and comprehension of technical texts, exposing the acquired knowledge, describing, analyzing and solving problems in real applications.

Lifelong learning skills:

Learning an appropriate studying methodology, supported by text consultation and implementation of the techniques proposed during the course.

Course program

- Singular value decomposition of a data matrix and Vector Space Model
- Data analysis by SVD factorization of the covariance matrix: SVD and Principal Component Analysis.
- Clustering and main clustering algorithms: k-means and k-medoids, hierarchical clustering algorithms

- Factorization of matrices (Nonnegative Matrix Factorization) and Clustering.
- Numerical classification algorithms, the Nearest Neighbors method, Learning Vector Quantization algorithm, Support Vector Machines.
- Analysis of HITS and PageRank methods for ranking complex networks.
- Numerical algorithms for problems on networks and graphs: Dijkstra's Algorithm (for the minimum cost path problem), Ford and Fulkerson's Algorithm (for the maximum flow problem).

Teaching methods:
Lectures session and exercises in computer lab

Auxiliary teaching:
Notes available at the web page https://sites.google.com/site/nicolettadelbuono/

Assessment methods:
Oral examination (on a given assignment)

Bibliography:
V. Comincioli, Metodi numerici e statistici per le scienze applicate, Milano, Ambrosiana, 1992.
- C. Meyer, Matrix Analysis and Applied Linear Algebra, SIAM, 2003.
- I.T. Jolliffe, Principal Component Analysis, Second Edition, Springer, 2002
- A. Cichocki, R. Zdunek, A.H. Phan, S.I Amari, Nonnegative Matrix and Tensor Factorizations, Wiley, 2009
- A. N. Langville, C. D. Meyer: Google's PageRank and beyond. Princeton Univ. Press, 2006.
- T. Hastie, R. Tibshirani J. Friedman: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition, 2009