Title: Low-rank approaches for data analysis: models, numerical methods and applications

Lecturer: Dr. Flavia Esposito, Department of Mathematics Università degli Studi di Bari Aldo Moro (flavia.esposito@uniba.it, https://sites.google.com/view/midasresearchgroup/people/flavia-esposito)

Abstract:

Low-rank approximation techniques represent an important branch of data analysis. Their use includes different scientific fields varying from preprocessing (treatment of missing values, anomaly detection, dimensionality reduction) to the proper knowledge extraction from the data.

Low rank dimensionality reduction mechanisms are recognized in the literature due to their ability to exploit the matrix (and sometimes tensor) structure of the data under consideration allowing to analyse both features and observations of the data. Low-rank approximation techniques are based on various mathematical tools ranging from linear and multilinear algebra to optimization.

It is therefore an interdisciplinary topic that touches on various fields, such as mathematics, computer science, information theory and the signal processing.

This course aims to present and examine some of the matrix and tensor decomposition models devoted to the extraction of knowledge from large, structured data. Known in the literature are the Singular Value Decomposition (SVD) and Principal Component Analysis (PCA), Independent Component Analysis (ICA), Nonnegative Matrix Factorizations (NMF) and all their variants in the tensor case (CP, HOSVD, HOOI, NCP, HONMF, TRI-ONMF, etc.).

The models under consideration almost always take the form of an optimization problem of a specific fitting function. Appropriate numerical methods for solving these optimization problems will be explained in detail in the course with the corresponding numerical algorithms.

Large part of the course will be also devoted to present and discuss several applications and real case studies. The students will be guided in the choice of the best algorithm in relation to the characteristics of the data under analysis.

Program:

The course will review some of the most famous low-rank models according to the matrix and tensor structures of the data. For these models, some numerical algorithms solving the associated optimization problem will be explained in details. Practical examples of real case applications to the biomedical field, text-mining, and satellite data analysis will be also presented.

Exam:

The final assessment will consist of the application of one technique studied during the course on a real dataset. Students can either attend the exam (the last day of the course) or offline at their convenience.

Participation:

Students willing to attend the course are encouraged to email the lecturer. A Google form will be sent to them to assess the background of the class.

Schedule and duration:

The course will be held presence: 10hours during the second week of January 2023 (11,12,13 of January 2023).