Computational linear algebra for large scale problems

Course Description

This course aims at presenting the mathematical and numerical foundation of several methods applied in Data Science. Analysis of large scale data sets requires specific algebraic tools in order to extract the most relevant information from data. This problem is tackled by the application of several mathematical tools; this course is designed to present them and to explain pro and cons of their application to realistic data sets.

Prerequisites

Basic knowledge of basic linear algebra and calculus is a prerequisite, as well as a basic coding ability and computer knowledge.

Course topics

1. *Basic linear algebra tools: vector spaces, bases, linear operators, matrices, eigenvalues and eigenvectors, norms.*
2. *Dense and sparse matrices, matrix operations on several computer architectures and performances analysis on CPUs and GPUs.*
3. Vector rotations, orthogonalization, projections: Grahm-Schmidt, Givens, Householder methods and QR factorization.
4. *Iterative solutions of large scale linear systems: applicability, convergence, computational cost and memory requirements, preconditioning.*
5. *Approximation of data and functions: global and piecewise interpolation, least square approximation, numerical tools.*
6. Eigenvalues and eigenvectors computations: numerical methods and common tools for large scale matrices. Stability and conditioning.
7. Computation and theoretical properties of Singular Value Decomposition (SVD).
8. Generalized inverse matrix and Moore–Penrose inverse.
9. Dimensional reduction of a problem and Principal Component Analisys (PCA).
10. Randomized SVD and Johnson-Lindenstrauss theorem.
11. Spectral Clustering

Course structure

Theoretical lectures and practice classes. Theoretical lectures are devoted to the presentation of the topics, with definitions, properties, introductory examples. The practice classes are devoted to train the students' abilities to solve problems and exercises and to perform computations and simulations with common tools.

Reading materials

Slides presented during lesson will be made available through the *Portale della Didattica*.
Other material will be suggested in class and, if possible, made available through the Portale della Didattica.
Suggested textbook:

- Linear Algebra and Learning from Data, G. Strang, Cambridge University Press, 2019, ISBN: 9780692196380

- Iterative Methods for Sparse Linear Systems, Y. Saad, Society for Industrial and Applied Mathematics Philadelphia, PA, USA, 2003, ISBN:0898715342

Assessment and grading criteria

Exam: compulsory oral exam;

Homeworks **HW1** and **HW2** consist of exercises that are aimed to evaluate the students in using the methods presented. They will be assigned to the students during the course

HW3 is **optional** and it is an application of the methods learned to a problem **chosen by the student**.

The oral exam will then consist of two parts:

a) a **discussion** of the submitted **HW1, HW2,** and optional HW3 **reports**, aimed at testing the depth of the students' understanding of the subjects and their ability to explain, defend, reflect, critically evaluate, and possibly improve their work, proving the real acquisition of the abilities listed in the expected learning outcomes section.
b) **a question about a topic studied in the course** covering both theoretical aspects and possibly their implementation and applications proving the real acquisition of the knowledge listed in the expected learning outcomes section.

Grading:

- The maximum grade for **HW1, HW2,** and HW3, upon the discussion detailed at point (a) above, is of **14 points**.
- The maximum grade for part (b) of the **oral test** is **18 points**.

The final course grade is then obtained by summing up the final grades of part (a) and (b) of the oral test.

Regrettably, the descriptions for the homework assignments will be provided towards the conclusion of the course, as they necessitate a comprehensive understanding of nearly all the topics covered throughout the course.

The submissions, which encompass both reports and code, must be uploaded to the "Elaborati" section on the course's webpage before the deadline for booking the exm in the chosen examination call (appello).

Furthermore, the homework assignments are encouraged to be undertaken in small groups comprising two individuals. However, should a student prefer to work on the homework individually, that option is also acceptable.

The students should upload a file for each HW with the following name-structure:

Surname1_Surname2_HW[1, 2, 3].zip

Throughout the course, you can expect presentations and discussions of Python and Matlab codes. Specifically, for the homework assignment concerning Principal Component Analysis (PCA), it is expected to be developed using Python. However, for the other homework assignments, you have the flexibility to choose between Matlab or, if you prefer, Julia as your programming language of choice.