

Mathematical Models and Numerical Methods for Omics Data Analysis.

Teacher

Flavia Esposito (<https://sites.google.com/view/midasresearchgroup/people/flavia-esposito?pli=1>)

Course description (min 150, max 300 words)

Interacting molecules and macromolecules drive complex biological processes. High-throughput technologies now allow us to collect omics data on the molecular components of biological systems, stored in mathematical structures such as nonnegative matrices or tensors. However, the high dimensionality, variable types, and noise level of this data pose significant challenges for multi-omics analysis. Addressing these challenges requires the development of novel computational methods to fully leverage multi-omics technology. This course aims to show mathematical models and numerical methods for omics data analysis based on low-rank dimensionality reduction approaches, in particular nonnegative matrix and tensor factorizations. The course is devoted to present the factorizations methods such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), Nonnegative Matrix Factorizations (NMF), their variants in the tensor case (Multilinear PCA/ICA, HONMF, TRI-ONMF, etc.) and their compared performance to omics data analysis purposes.

Low-rank dimensionality reduction mechanisms are recognized in the literature due to their ability to exploit the matrix/tensor structure of the data under consideration allowing the analysis of 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. Their use includes different scientific fields from preprocessing (treatment of missing values, anomaly detection, etc.) to the proper knowledge extraction from the data. Since the nonnegativity nature of the omics data, from a mathematical point of view, it is convenient to formulate omics data analysis tasks as constrained optimization problems of specific divergence functions. Appropriate numerical methods for solving these optimization problems will be explained in detail in the course together with the ad-hoc developed numerical algorithms.

Course period

September-October 2024

SSD

MAT/08

Course References (optional)

- Gillis 2020. *Nonnegative matrix factorization*. Society for Industrial and Applied Mathematics, 2020.
- Esposito et al 2020. An NMF-Based Methodology for Selecting Biomarkers in the Landscape of Genes of Heterogeneous Cancer-Associated Fibroblast Populations. *Bioinform Biol Insights* 8

- Forcato et al 2021. Computational methods for the integrative analysis of single-cell data, *Brief Bioinform* 22(1):20-29
- Lê Cao et al 2021. *Multivariate Data Integration Using R: Methods and Applications with the mixOmics package*. CRC Chapman & Hall.
- Liu et al 2016. Block-Constraint robust principal component analysis and its application to integrated analysis of TCGA data. *IEEE Trans Nanobiosc* 15(6):510–6.
- Meng et al 2016. Dimension reduction techniques for the integrative analysis of multi-omics data, *Brief Bioinform* 17(4):628–641
- Rohart et al 2017 *mixOmics: an R package for omics feature selection and multiple data integration*. *PLoS Comp Biol* 13(11):e1005752
- Selicato et al. 2021. A New Ensemble Method for Detecting Anomalies in Gene Expression Matrices. *Mathematics* 9:882.
- Stein-O'Brien et al 2018. Enter the matrix: factorization uncovers knowledge from omics. *Trends Genet*.
- Zhang and Zang 2019. Learning common and specific patterns from data of multiple interrelated biological scenarios with matrix factorization, *Nucleic Acids Res* 47(13):6606–6617

Credits and Hours

2 credits (16 hours)

Exam Modality

Two alternatives are available to the student to pass this exam (Teacher may choose other modalities):

- 1) Paper presentation. Students present the content of 2 papers suggested by the teacher. No groups are allowed.
- 2) Project. Students implement and experimentally validate an algorithm or its variation from a paper suggested by the teacher. Projects can be done in groups of 1-3 students, depending on the algorithm.

Teacher CV

Attach or link a max 3 pages CV for each teacher proposing the course.

Teacher Main Publications

- Bi-level algorithm for optimizing hyperparameters in penalized nonnegative matrix factorization. N.Del Buono, F.Esposito, L.Selicato, R.Zdunek. *Applied Mathematics and Computation* 457, 128184, 2023
- Theoretical aspects in penalty hyperparameters optimization. F. Esposito, L. Selicato, C. Sportelli. *Mediterranean Journal of Mathematics* 20.6, 300, 2023.
- Cluster of resistance-inducing genes in MCF-7 cells by estrogen, insulin, methotrexate and tamoxifen extracted via NMF, A.Boccarelli, N.Del Buono, F.Esposito, *Pathology-Research and Practice* 242, 154347, 2023.
- Electronic case report forms generation from pathology reports by ARGO, automatic record generator for onco-hematology, GM.Zaccaria, V.Colella, S.Colucci, F.Clemente, F.Pavone, [], Flavia Esposito, et. Al, *Scientific Reports* 11(1), 23823, 2022
- Colorectal cancer in Crohn's disease evaluated with genes belonging to fibroblasts of the intestinal mucosa selected by NMF, A. Boccarelli, N. Del Buono, F. Esposito. *Pathology-Research and Practice* 229 (2022): 153728.

- A review on initialization methods for Nonnegative Matrix Factorization: towards omics data experiments, F.Esposito, Mathematics, 2021.
- A new ensemble method for detecting anomalies in Gene Expression Matrices, L.Selicato,F.Esposito,G.Gargano,MC.Vegliante,G.Opinto,GM.Zaccaria,S.Ciavarella,A.Guarni,N.Del Buono, Mathematics, 2021.
- A NMF based methodology for selecting biomarkers in the landscape of genes of heterogeneous cancer associated fibroblast populations, F. Esposito,A.Boccarelli,N.Del Buono, Bioinformatics and Biology insights,volume 14, 2020
- Orthogonal Joint Sparse NMF for microarray data analysis, F.Esposito,N.Gillis,N.Del Buono, Journal of Mathematical Biology, 2019
- Improving knowledge on the activation of bone marrow fibroblasts in MGUS and MM disease through the automatic extraction of genes via a nonnegative matrix factorization approach on gene expression profiles, A.Boccarelli,F.Esposito,N.Del Buono,A.Vacca,M.Coluccia, Journal of Translational Medicine Journal of translational medicine, 16(1):217, 2018