
M2 internship – Longitudinal data analysis in high dimension

Description

In collaboration with clinicians and biologists from various teams of Lille, our METRICS team has developed an expertise on high-throughput analysis of omics data (e.g., genomics, transcriptomics, proteomics). These data present much more variables than individuals and penalised regression framework has proven to be very useful to select biomarkers at one time point. When several time points are considered, interpretation of biological results is much more difficult if time points have been studied independently.

The aim of this internship is to apply, on real data, methods which are conceptually different for the analysis of longitudinal data in high dimension :

- rJLCM - regularized latent class model (Sun et al., 2019)
- multiway generalized canonical correlation analysis (A. Gloaguen et al., 2022)

Missions

- Analysing real data using the R code of the authors associated to previous publications
- Performing a bibliographic review to suggest other analyses to select markers in a longitudinal study of high-throughput experiments

Necessary skills

- Advanced R programming
- Statistical learning in a context of high dimension

Skills to be acquired

- Fluenced R programming
- Survival analysis
- Multi-way analysis
- Joint models

Practical information

Duration : 6 months (beginning depending on the master)

Location : ULR2694 METRICS – Pôle recherche de la faculté de médecine – 1 place de Verdun, 59000 Lille (FRANCE) or Inria MODAL – 40 avenue Halley, 59650 Villeneuve d’Ascq (FRANCE)

Contact

Génia Babykina evgeniya.babykina@univ-lille.fr and Guillemette Marot guillemette.marot@univ-lille.fr

Bibliography

Gloaguen A, Philippe C, Frouin V, Gennari G, Dehaene-Lambertz G, Le Brusquet L, Tenenhaus A. Multiway generalized canonical correlation analysis. *Biostatistics*. 2022 Jan 13;23(1):240-256. doi: 10.1093/biostatistics/kxaa010. PMID: 32451525.

Sun J, Herazo-Maya JD, Molyneaux PL, Maher TM, Kaminski N, Zhao H. Regularized Latent Class Model for Joint Analysis of High-Dimensional Longitudinal Biomarkers and a Time-to-Event Outcome. *Biometrics*. 2019 Mar;75(1):69-77. doi: 10.1111/biom.12964. Epub 2018 Dec 5. PMID: 30178494.