

Neural Networks for Transcriptomic Data Analysis

Projects

Aims

- Apply the techniques covered in class to analyze and interpret transcriptomic data from TCGA subsets.
- Build and test MLPs (supervised) and autoencoders (unsupervised); tune hyperparameters systematically and log runs.
- Evaluate models with appropriate metrics (e.g., ROC-AUC/F1 for classification and MSE/MAE for regression) and report uncertainty.
- Identify important variables/features and provide a biological interpretation of the results

Methods and groups

- Methods
 - Data prep: QC, normalization/standardization, train/val/test with fixed seeds.
 - Start simple and iterate: begin with a minimal MLP, then extend architectures; add autoencoders for representation.
 - Hyperparameter tuning: architecture depth/width, learning rate, regularization, early stopping; keep a structured run log.
 - Evaluate with clear metrics and document variance (CV or repeated runs) before moving to interpretation.
- Project groups :
 - Groupe of 3 students
 - Mixed by BMC/BIP tracks
 - Groups different from other project groups

Biological Questions

- You have complex datasets on which you wan ask multiples question:
 - Diseases vs. control
 - High-grade vs low-grade
 - Treated vs untreated
 - ...
- You are free to ask any questions that you want
- Keep in mind :
 - You need a well balance between methodological questions (architectures, metrics) and biological interpretations (gene, function, pathways)

Deliverables

- Paper (~15 pages):
 - context, data & preprocessing, methods, results, biological interpretation, limits & improvements
 - Report expected for January 11th
- Code:
 - clean git hun repo, README with commands, environment file, data access notes, scripts/notebooks documented.
 - Code expected for January 11th
- Oral (20 minutes + 30 minutes of Q&A):
 - clear story, fair speaking time, results and interpretation, defend choices & limitations.
 - Slide expected for January 11th
 - Presentations will be made during the week of the 12th of January

Everything must recreate reported results without manual tweaks