Course Project

Option 1: ML analysis of Human Embryonic Cells

Data & Background

Gene expression data of selected genes of cells in human embryonic cells of day 3 to day 7, labeled as E3, E4, E5, E6 and E7. The data were selected from [Petropoulos et al. Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. Cell, 2016, 165(4): 1012-1026]. Each cell is represented as a vector of the expression of 490 genes.

Task 1. Mapping clusters between days & Mapping clusters to known lineages

Do unsupervised learning on E5 and E7 cells respectively to identify clusters in each day, mapping the clusters between days, and mapping the clusters with known lineages.

Note: There are four known lineages in human early embryonic cells, they are: trophectoderm (TE), epiblast (EPI), primitive endoderm (PE), and pre-lineage. You should read the original paper to find the meaning and pattern of these lineages.

Task 2. Classifying cells into lineages

Use E5 clusters (lineages) to train a lineage classifier (with feature selection), use CV to study the performance. And then:

2.1 Apply the classifier on E6 and E7 cells and analyze the results.

Note: It is expected that cells in E6 and E7 are well differentiated, so they probably will not be classified as pre-lineage.

2.2 Apply the classifier on E3 cells.

Note: We expect that cells in E3 are undifferentiated, so they will all be classified as pre-lineage.

2.3 Apply the classifier on E4 cells and analyze the results.

Note: This is the point where new discoveries may be expected.

Option 2: Self-Proposed Projects

You can also choose another project by yourselves. It must meet the following requirements:

1. It must be a real research project with real questions (not a toy study).
2. It must involve multiple (≥3) methods of different categories learned in this course.
3. You must already have the data and facility and background for the proposed project. If it is a real project from some professor's lab, you need make sure that the professor agrees the project to be a collaborative project with this course.
4. You need to write a proposal and get approved. Feedbacks will be given within 1 week after you submit your proposal. And you need to re-write a proposal if the previous one is not approved. You may need to be interviewed with the TAs or the Instructor to get the proposal approved.

Deadline for submitting the proposal for the project: Nov. 24 (Friday) 23:59, 2018
Final Report Due: Jan. 6 (Sunday) 23:59, 2019

Requirements on Final Project

The final project must be completed by teams of 2-3 students.

A proposal on your choice of the final project must be submitted by the deadline. Members of the same team should all submit the same proposal in the homework system. The proposal should be within one page, with at least the following information:

- Title of the project
- Names of team members (mark the team coordinator)
- Expected contribution from each team member

For teams working on Option 2, the proposal should also include the following extra information:

- Background of the project
- Origin of the project (How/where did you get this project?)
- Data availability
- Major questions to be studied in this project
- Expected possible results
- Key references (if any)

A formal technical report must be submitted by the due date. The report should be either in English or in Chinese with proper formatting similar to the style of formal journal publications.

The report should be no more than 6 pages including figures, tables and full list of references (A4 size page, with font of main text of English 10~11pt or Chinese 五号字). Reports should be submitted in PDF formats only. Supplementary files should be used to include necessary technical details if they cannot be included in the main text. The supplementary files should include software codes and documents, necessary intermediate data, results, etc. Supplementary files should be submitted as .zip or .rar files.

The contribution of each members in the team should be clearly stated in the report. Any discussions with anyone else of this class or out of this class should also be acknowledged in the report.