

IE 8623 Course Project Description

Fall 2021

As a part of the course assignments, you are required to complete a project as described below.

Objective: Conduct the regression and/or classification analysis introduced in class, and compare their performance in terms of accuracy and computation time. Interpret your results.

Dataset: You should find a data set with dimension greater than 5 (functional data with lower dimension would also be fine). You can obtain data from the following three sources

- 1) Research from your own lab with permission of your advisor;
- 2) The following link: <https://archive.ics.uci.edu/ml/datasets.php>;
- 3) Contact the instructor for other available data sets.

Software: R, Python, Matlab, SAS, or any other tools you feel comfortable to use.

Project Group: each individual student forms one team.

Project Grading: proposal (20%) + final report (80%)

Proposal due: 10/31/2021.

In the proposal, please answer the following questions.

1. Briefly describe your data set
 - a. What are the input/output variables?
 - b. What are the data formats of those variables?
 - c. Are there any missing data?
 - d. What analysis task(s) do you plan to perform (e.g., regression, classification, etc.)?
2. What are the algorithms you plan to use (linear discriminant analysis, k nearest neighbors, naïve Bayes classifier, support vector machine, etc.)? You are expected to use at least three methods we covered in class.
3. Visualize your data properly;
4. Clarify what is the objective of your project (e.g. input-output relationship you want to explore, optimize the output performance, etc.)

* If you are using a dataset from your own lab, please specify what's new you would like to explore.

Final report due: 12/8/2021.

The final report is expected to have the following sections:

1. **Dataset description**¹
2. **Data visualization**
3. **Methods and performance comparison:** The grading will not be based on your reported accuracy, but if the methods are applied in a reasonable way. You are expected to use *at least three methods* we covered in class.
4. **Possible improvements:** State potential approaches that can be helpful to improve the performance of your models.
5. **Conclusion**

¹ You may directly use your approved proposal information for Section 1 & 2