

## Final Project for MGSC 695: Optimization for Data Science

For the final project for this course, you are required to implement a machine learning method of your choice from scratch and test its performance on an appropriate dataset. That is, you cannot use a black box package or module of the method. However, you are also expected to compare the performance of your implementation with any available black-box implementation of the same method. Your project implementation will be judged on its difficulty and the extent of execution. Each group member will be awarded identical points for the project. The due date for the final project is **March 5th, 6 pm**.

Along with the code, you are also expected to submit a project report.

The report will summarize the steps involved in the project. Ideally, this report will contain

- (i) description and details on the method thereby exhibiting your understanding of the method you have chosen,
- (ii) (details on the underlying optimization problem and the approach to solve it (the formulation and identify whether it's a linear or a quadratic or a convex optimization problem etc., the optimization approach and so on).
- (iii) You can also use the report to explain the challenges you faced and whether you were successful in overcoming them.

Please always cite any resources you have relied on. The report should be no longer than 10 pages (at least single line spacing and 12 pt font). **Your project report will be evaluated by the depth of your understanding, clarity, presentation, and completeness. A list of potential topics is provided below, however, you are free to pick any other method as well.** If you are deviating from the list, please confirm with me. You are also free to demonstrate your implementation on any appropriate dataset.

1. Mixture Models (GMM or other mixture models)
2. K-means clustering (related methods and extensions)
3. Regularized regression methods (Lasso, ridge regression, etc.)
4. Separating hyperplanes (Reference: Section 4.5 in Elements of Statistical Learning by Gareth James et al.)
5. Tree-based methods (e.g., regression trees, decision trees)
6. Association learning methods (e.g., Apriori, ECLAT)
7. Hierarchical clustering methods
8. Dimensionality reduction (e.g., Principal Component Analysis)
9. Support Vector Machines

## Groups for Final Project

### Group 1

Agarwal, Lakshya
Anwar, Arham
Balakrishnan, Jared Arvind
Barabasz, Michelle

### Group 2

Batbayar, Oyundari
Belmadani, Samia
Cai, Wenya
Castro Canal, Rodrigo

### Group 3

Chaturvedi, Jaya
Clarke, Sean
Delisle, Audrey
Elenany, Mohamed

### Group 4

Gao, David
Gao, Ge
Gupta, Krishan
Huang, Yanhuan

### Group 5

Jackson, Reo Paul
Ji, Addison
Kao, Kelly
Lagarde, Tiffany

### Group 6

Liew, Chien
Liu, Jennifer
Lou, Ivy
Lu, Yifan

### Group 7

Lu, Yu
Luo, Xingchen
Najejan, Valentin
Park, Seunghyun
Quansah, Jessica

### Group 8

Sangwan, Om
Song, Jessica
Wang, Jiaxuan
Wang, Ko-Jen
Wang, Siqi

### Group 9

Yadav, Nandani
Yang, Yiyi
Yu, Yichen
Zhong, Zhicheng
Zhu, Yvette