Conclusion

In this project, we performed a data analysis cycle which contains getting raw datasets, data cleaning and processment, exploratory data analysis, modeling and tuning parameters, visualization of results, report and presentation.

We replicated major findings on chapter 6: Linear Model Selection and Regularization (from "An Introduction to Statistical Learning" by James et al) with the dataset "Credit". we performed five algorithms: Least Square Regression, Ridge Regression, Lasso Regression, Principal Components Regression, and Partial Least Squares Regression.

From our result, we found that Lasso Regression performs best, followed by Ridge Regression using the measure of test Mean Squared Error. Dimension Reduction Regression, in this case, do not perform as well as Shrinkage Methods. The three leading predictors in prediction are Limit, Income and Rating.

References Lecture slides of Stat 159

"An Introduction to Statistical Learning" by James et al