Introduction

The goal of this project is to perform alternate regression analyses to find the best model that produces the best response variable (balance).

In this project, we will look at our data set credit and perform different regression analyses. We will first pre-process the data set to set up variables to our likings.
We then shall standardize the data set and then explore credit to look for correlations between variables.
From that, we can determine which variables are closely related and which are not related at all to each other and to our response variable, balance.
We then shall look at which categorical variable has a significant effect on the response variable.

Once we have thoroughly analyzed the data set properties and variables, we can go ahead and run our 5 picked alternative regression models: OLS(Ordinary Least Squares), RR(ridge regression), LR(lasso regression), PCR(principle components regression) and PLSR(partial least squares regression).

By looking at the MSE(mean squared error) of which each model produced, we can determine which model generated the best regression, and the best fit for our response variable.
We will explore all of these concepts in our project.