Introduction

Since we are predicting for a continuous output, we choose Least Square Regression as our base model. Then we choose two Shrinkage Methods (Ridge Regression and Lasso Regression) and two Dimension Reduction Methods (Principal Components Regression and Partial Least Squares Regression) to compare the results with our base model.

For parameter choosing, we use 10-fold Cross Validation to choose minimum lambda for Shrinkage Methods and minimum validation components for Dimension Reduction Methods. Then we compare results by splitting dataset into training set and test set and calculating Mean Square Error on test dataset. Finally we calculate the coefficients in each of five methods used for prediction to see which factor matters most in prediction.