Predictive Modeling Process

Atul Lanka, Rushil Sheth

November 4th, 2016

Abstract

This project is based on the data set and statistical concepts introduced in Chapter 6, **Linear Model Selection and Regularization**, from "An Introduction to Statistical Learning" (by James et al.). The primary objective is to perform mulitple predictive modeling processes applied to the data set *Credit*, as well as learn about the relationship between `Balance` and the rest of the quantitative and qualitative variables. The five models considered are *Ordinary Least Squares*, *Ridge Regression (RR)*, *Lasso Regression (LR)*, *Principal Components Regression (PCR)*, and *Partial Least Squares REgression (PLSR)*. Using cross-validation, the most ideal model is evaluated from their respective Minimum Square Errors on the test sets.

Introduction

The primary purpose of this report is to determine the best model for predicting `Balance` given the ten different predictors, both quantitative and qualitative from the *Credit* data set. The distributions of these variables will be examined through summaries and plots, and five aforementioned regression models will be applied to the data (*Ridge Regression (RR)*, *Lasso Regression (LR)*, *Principal Components Regression (PCR)*, and *Partial Least Squares REgression (PLSR)*). The means of comparing and analyzing five models will be by studying their respective coefficients and mean squared errors (MSE) calculated from the 10-fold cross-validation.

We will discuss our data, methods, analysis, and our main conclusions throughout this paper. These sections will also include diagrams such as tables and graphs to help the reader gain a better understanding of our data and to visualize the outcomes of the different methods. data stuffff

methods stufffff analysis. . . .need to read the book for this results. . . what happened??

what can we take from this??