Coordinate Descent Algorithm for solving LASSO and Elastic Net Regressions

1. Summary

In this section we compare the performance of the LASSO and Elastic Net Regressions in prediction accuracy and variable selection. We run both by determining the penalization coefficients, for LASSO, and and for the Elastic net, through the Coordinate Descent Algorithm. First, we give a brief explanation of how the penalizations of LASSO and the Elastic Net work, and what effects each have. Then we explain how we implemented the Coordinate Descent Algorithm for each, followed by a description of how we finetuned the parameters. After that we show the testing results for each model and compare their prediction accuracy through their Mean Squared Error (MSE) and how they perform in variable selection through the number of regressors that are not equal to zero. The baseline results are be based on the parameters specified in the project instructions, which are indicated further below. In order to find out which model is more suitable depending on the situation, we compare their results with different parameters.

1. LASSO Penalization compared to Elastic Net

The LASSO regression is a model based on the linear regression that shrinks the regression coefficients depending on a parameter by minimizing the following expression:

Which is an extension of the linear regression, adding a penalty through to keep the size of the coefficients low, which is how it achieves the shrinkage. It is generally useful when dealing with datasets with a high number of predictors, as this penalization tends to shrink the coefficients easily to 0 when they are not relevant, as opposed to other shrinkage methods, like the Ridge Regression.

The Elastic Net is a combination of LASSO and the mentioned Ridge Regression. It tries to minimize the following expression:

The added penalization with is how the Ridge Regression achieves its shrinkage. Generally, Ridge tends to make more accurate predictions, but when there are many variables, LASSO outperforms due to its better variable selection. The Elastic Net tries to find a compromise by combining the two penalizations.

1. Coordinate Descent Algorithm

In order to determine the optimal coefficients for each model, we use the Coordinate Descent Algorithm. We programmed our own R function

Which does the following:

1. It initializes all the coefficients
2. Repeats the following steps until the values for all converge
   1. Computing of partial residuals:
   2. Computing the Simple Least Squares Coefficients of the residuals on each predictor:
   3. Update by soft thresholding:

This function is applicable to both the Elastic Net and the LASSO. In case we are using Elastic Net, it performs the soft thresholding with the specified and . In case we are doing LASSO, where we only have , the expression containing is equal to 1 and doesn’t affect the result.

We have defined the convergence condition for the coefficients as […what we have with difference in sum no larger than constant, or better something that considers each beta with RMSE]

1. Validate Models through Parameter Tuning

To choose the optimal lambdas for our models we programmed our own validation function, which iterated over all possible combinations of and (only in case of LASSO) we determined by calculating the optimal coefficients and a validation MSE for each. Once we have a validation MSE for each lambda combination, we choose the one with the lowest MSE

“After we have obtained optimal tuning parameters, we re-estimate the model on the training data and obtain our final results by running it on the test data.”