

PS9 Fretland

efretland

April 2018

1 Responses

Question 5

The dimension of the training data is 80 percent of the total rows in the dataset. The training data is a matrix of 404 cases with 450 variables each once the 6-fold validation is done.

Question 6

The in sample RMSE is 0.233. The out of sample RMSE is 0.203. Ideal lambda is 0.0875.

Question 7

The in sample RMSE is 0.177. The out of sample RMSE is 0.1452. Ideal lambda is 0.08.

Question 8

The in sample RMSE is 0.17. The out of sample RMSE is 0.1402. Ideal lambda is 0.0396 and ideal alpha is 0.0757. Because the ideal alpha value is so low, this is a strong argument for using a LASSO model.

Question 9

A linear model could not be used because of the cross validation that was done to the dataset. A linear model requires no regularization so therefore it cannot be applied to the cross-validated dataset.

Based on the comparisons to in and out of sample RMSE, The best model is the elastic net one as it has both the lowest in and out of sample RMSE. However, none of these make complete sense, as I expected the in sample RMSE to be consistently lower than the out of sample RMSE.

Since the out of sample RMSE is lower, it seems as though we are too far on the variance side of the bias variance tradeoff. We clearly are not too biased, as there is probably a good amount of underfitting going on.