Exam 1

Brian Lois

2/16/2017

1. Load the dataset into R.
2. Center and scale numerical predictors.
3. Create dummy variables for any categorical predictors.
4. Split the data into a training and test set. Set aside the test set until the end.
5. Split the training data using 4 fold cross validation.
6. Fit ridge regression models for a range of λ_2 values. Be sure to include large enough values of λ_2 that you see a decrease in performance.
7. For each value of λ_2, you will have 4 models (1 for each fold). Evaluate the RMSE of all models on the fold not used to train. Use a loop for this.
8. Make a plot with λ_2 on the x-axis and the mean RMSE (average over the 4 folds) on the y-axis.
9. Using this plot, select λ_2 for your model. Explain your reasoning.
10. Fit a model on the complete training data using your selected value for λ_2.
11. Evaluate the R^2 and RMSE of your model on the test set.