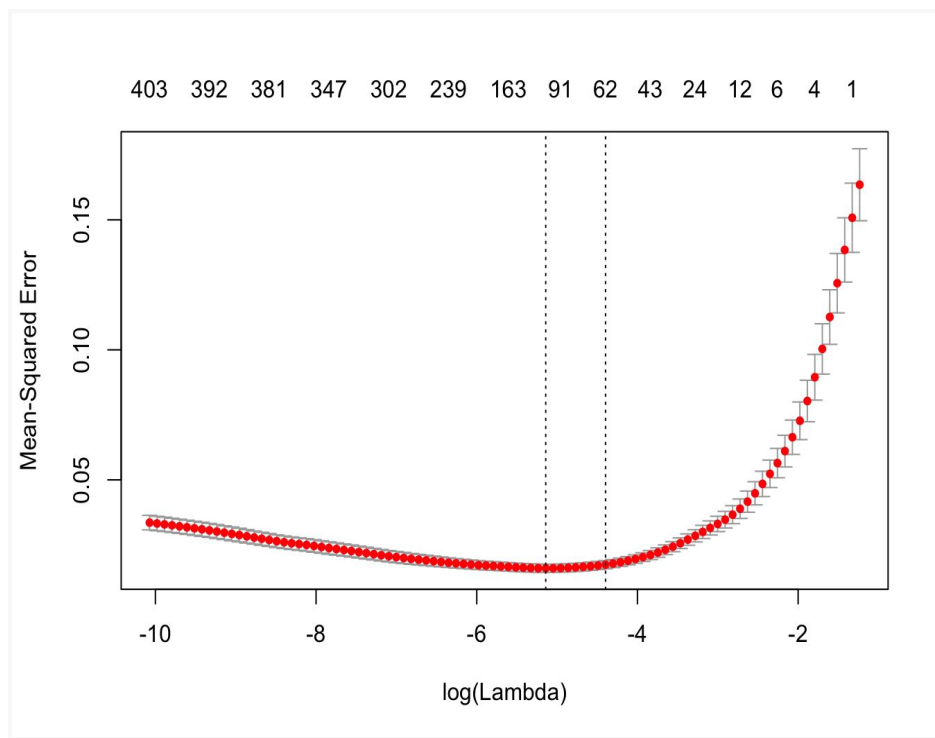


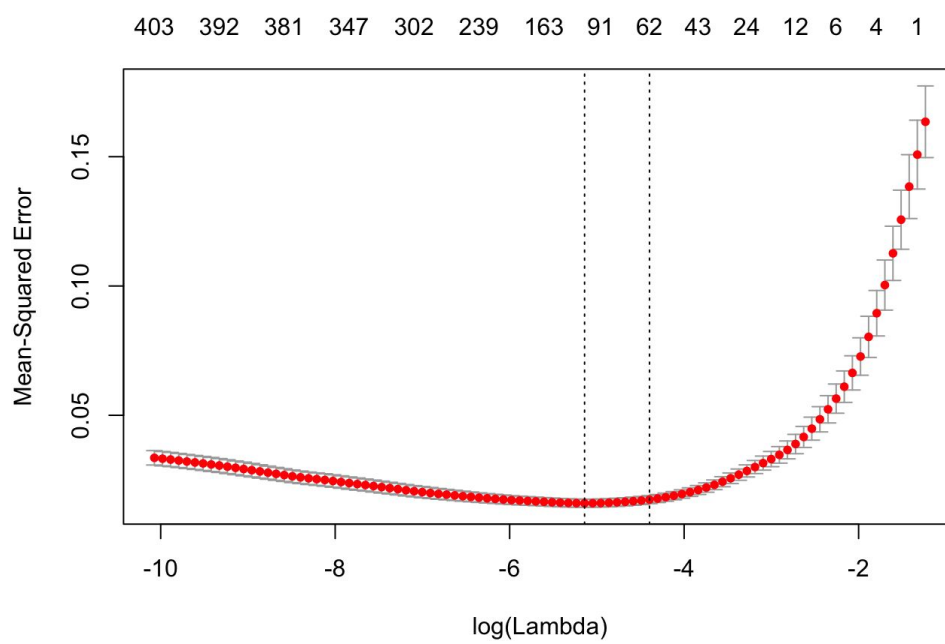
Besides Random Forest, which produced nice RMSE result but clearly not the best, I tried Ridge and Lasso to see if they could produce better predictions.

First of all, I did stepwise regression to predict sale prices, with forward selection as well as backward selection, but the RMSE was even worse than that of random forest, so I decided to give up on stepwise.

Second, I ran the following procedures of Ridge and Lasso with original data but the RMSEs were too large, so log transformation was applied to the data. Then I turned to Ridge and Lasso, all qualitative variables were turned into dummy variables, and the dataset was splitted into training and testing, glmnet package was applied to perform ridge and lasso, lambda was defined between 10^{-2} and 10^{10} , and then the best value of lambda that would minimize the error was chosen, in this case, cv.glmnet was applied to find out the minimum error, in which procedure the two regressions bridges the gap differently, lasso ($\alpha=1$, the default) and ridge ($\alpha=0$). Two plots of mean-squared error were created for ridge(1), and then lasso(2), one can see the cross-validation curve (red dotted line), with upper and lower standard deviation curves. And finally, ridge model with best lambda was tested, which produced an RMSE of approximately 0.016, and lasso produced an RMSE of around 0.017.



(1)



(2)