Capstone Lending Club

Mingyang Ke

Github: alvinke12

https://github.com/alvinke12/Capstone-Lending-Club/blob/master/r%20to%20now.R

Goal: use R to predict whether a borrower will miss payment, loan status of fully paid and current are classified as 'okay' and the rest as 'past due', which is also the target variable. Both logistic regression with lasso regularization and random forest were used to build model and compare ROC area. The first quantile squared error was used to select best lambda for lasso as shown in figure 1. The ROC curve area for logistic regression is 0.9714 shown in figure 2. Random forest seems to perform not as well as logistic regression and the ROC area is 0.8347 as shown in figure 3. Thus, the selected model for predicting whether a borrower will miss payment is logistic regression model built in r

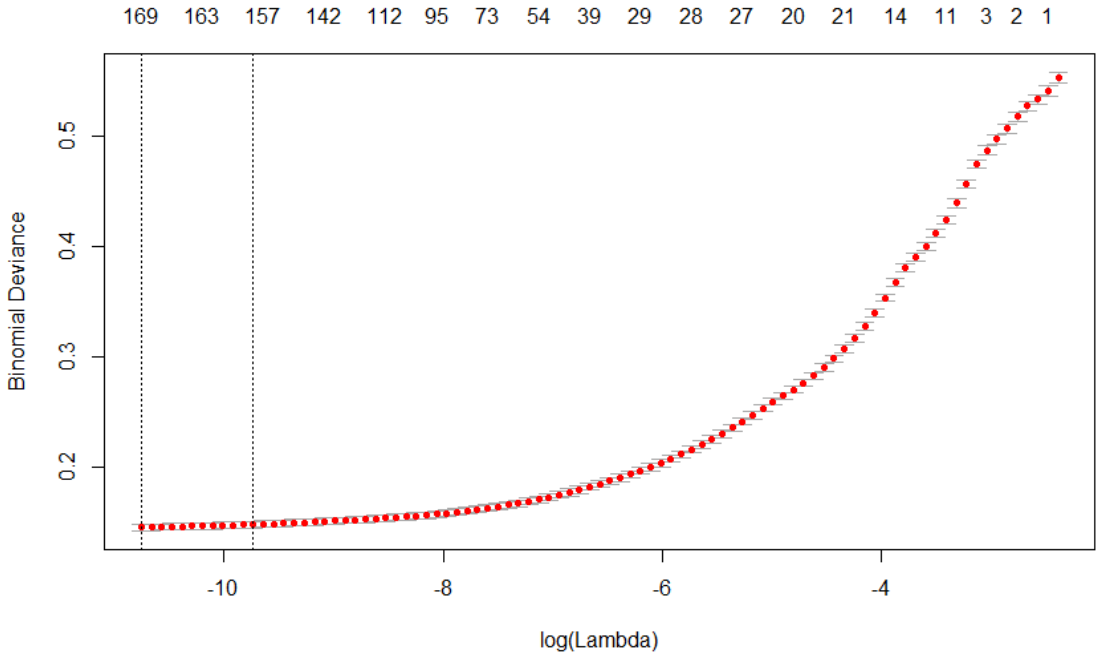

Fig. 1 1se for selecting best lambda

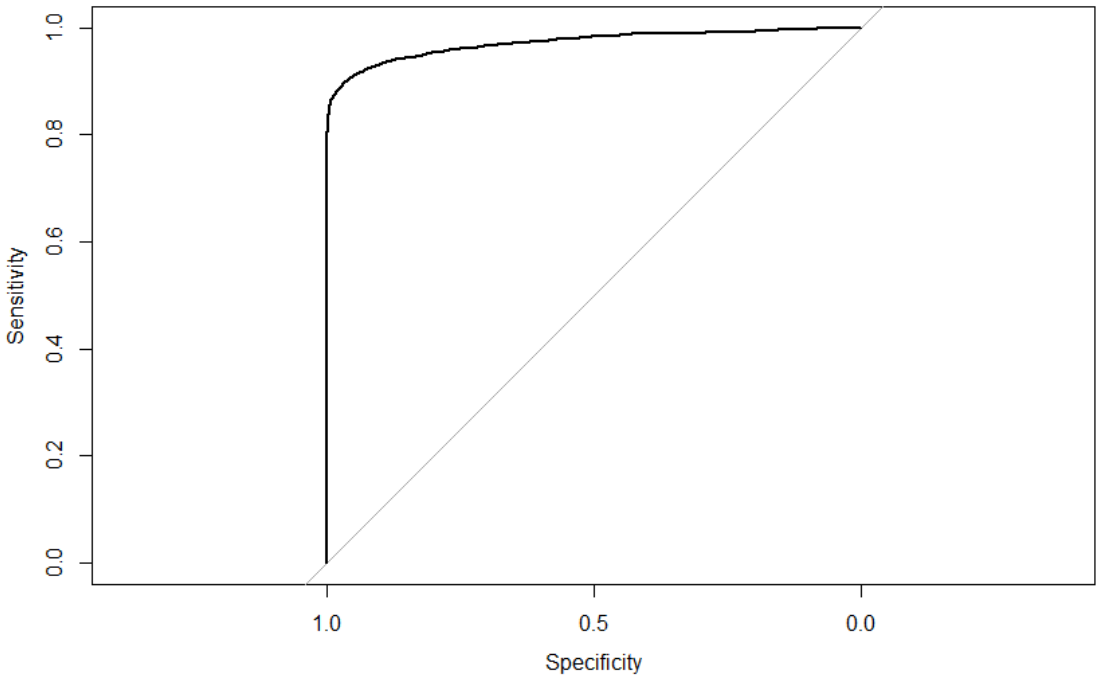

Fig. 2 ROC curve for logistic regression

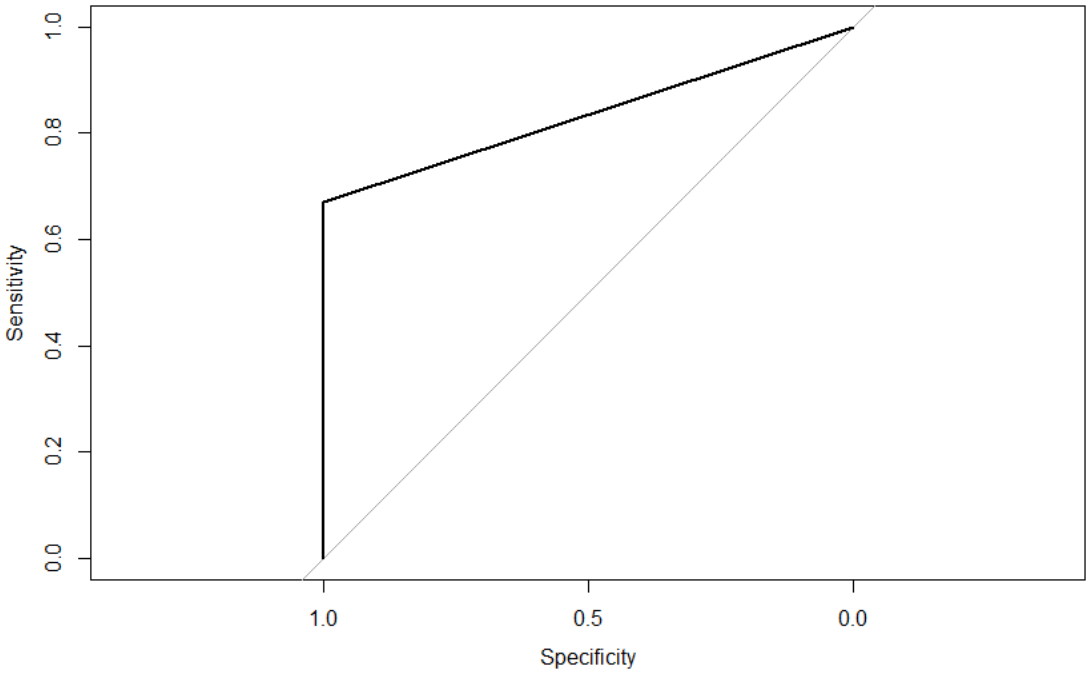

Fig. 3 ROC curve for random forest