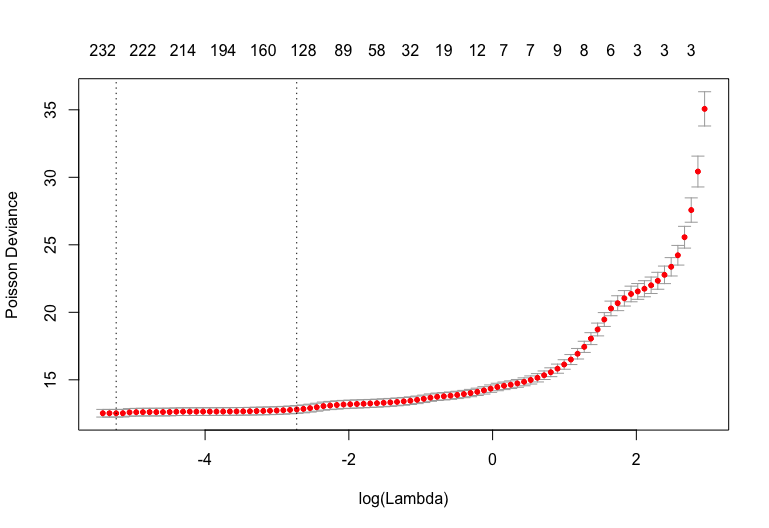
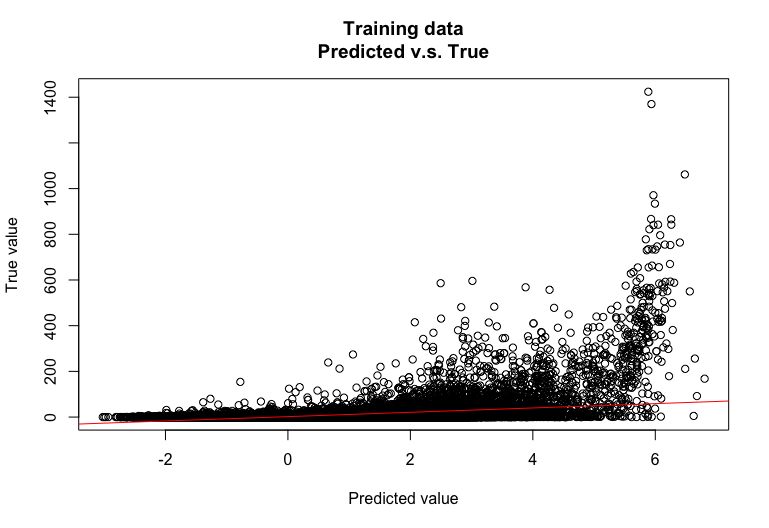
Problem 12.3 Yu Che Wang/ yuchecw2

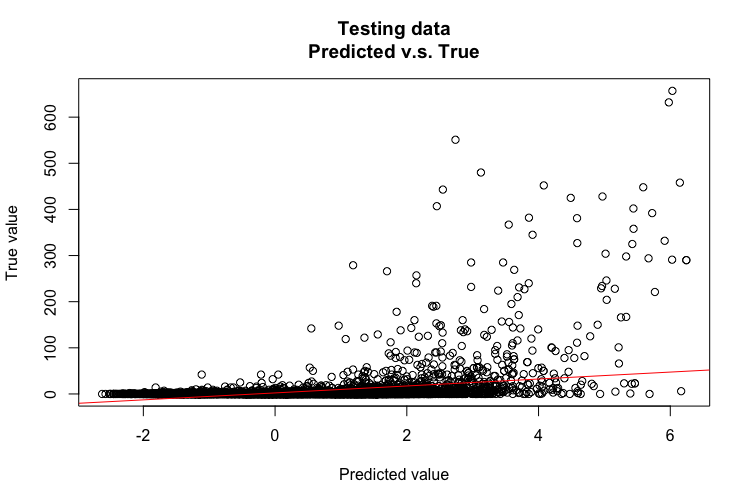
1. Plot of the cross-validated deviance of the model against the regularization variable



1. We chose the regularization variable to be lambda.1se=0.0789. We chose lambda=lambda.1se instead of lambda.min to prevent overfitting the training data.



1. We use the model which was trained on the training data to predict testing data. We chose lambda=0.0789.



1. Training dataset has larger standard deviation than testing dataset. The training dataset has more outliers. Both training and testing data have a large portion of data with zero y value (number of comments) which causes this model have low r squared value. Hence, this regression is difficult.

Code:

