Group 13

SSC 442 – Lab 03
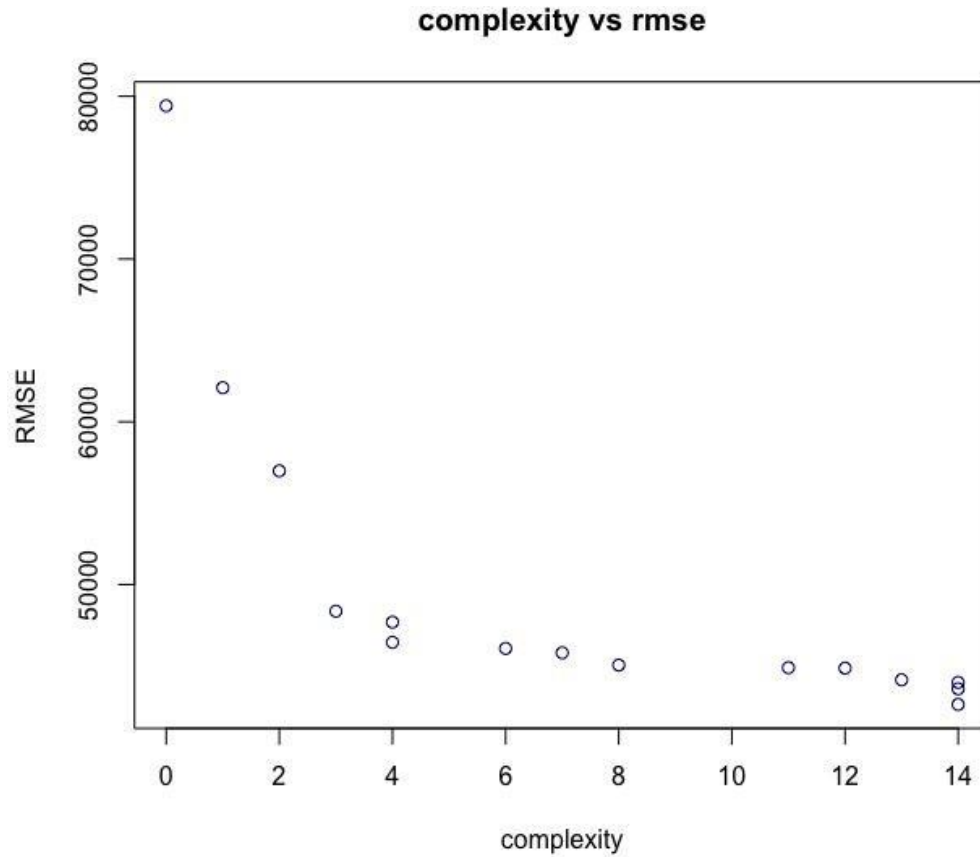
2/13/2020

<div align="center">Lab3</div>

**EXERCISE 1:**

1. Drop OverallCond and OverallQual

   Coding part: *Ames <- select(ameslist, -c(OverallQual,OverallCond))*

2. We created a series of models up to complexity length 15. The data is too complex and

   takes a long time, which is the disadvantage of a full-size model. Due to this, we do not

   believe it makes sense to use a full-size model.

EXERCISE 2:

The visualization we created has a couple of functions to it. First, we have the data that is

attempting to fit a model (train_rmse). This doesn't tell any sort of specific story, so the

other part of this is the test_rmse variable which describes how well the train_rmse data and

model fit one another. With
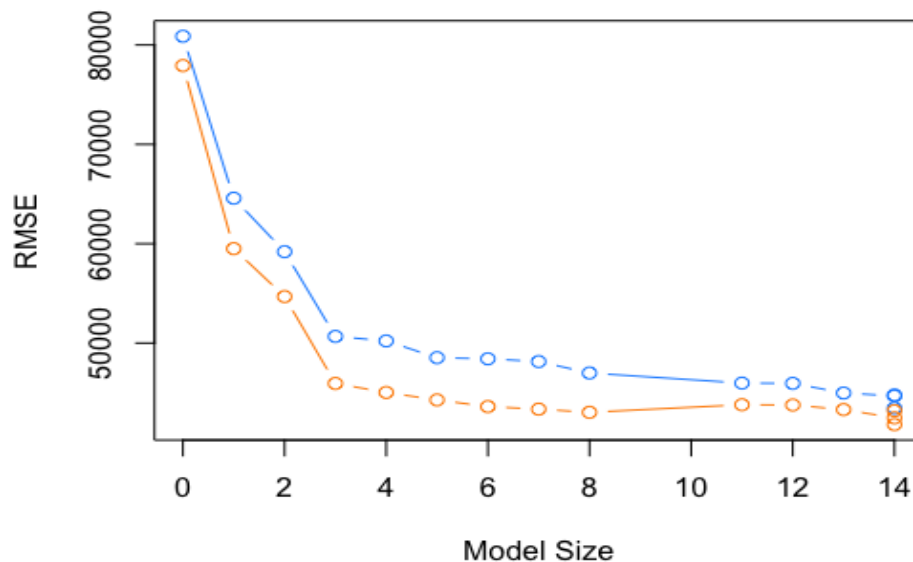
this, we can tell if

we are accurately

fitting and

predicting with our

models.

After taking five

different linear

models with

attributes such

as GrLivArea

(continuous), Heating (categorical), and compiling them to create the train_rmse model,

we were able to produce this graphic. (Note: all of the data not used by the train_rmse

variable was used in the test_rmse variable)

In regard to how well the accurately the train_rmse is able to predict, it can be seen that it is close to being deemed a good predictor. However, there are fitting issues. The blue line, representing the train_rmse model, looks almost as if it is inverse to the test_rmse model (orange line) in relativity to the data points on the opposite side of the graphic. If compared to other representations within the classroom, this may stand well on its own however different descriptors and clearer explanation of the visualization may supersede over ours.