## Question 1

What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

**Solution:**

The optimal value of alpha for ridge and lasso regression are 8 and 100 respectively as after that as the alpha keeps increasing the emphasis on the overall accuracy of the model goes down which means the model is going to make more and more error.

Hence choosing the double the value of alpha for both ridge and lasso will increase the error and accuracy of the model will deteriorate

## Question 2

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

**Solution:**

After finding the optimal value of lambda for both ridge and lasso regression, I would like to go with Lasso regression as a lot of coefficients seen at the end in case of lasso are zero, which means that lasso has performed variable selection.

Lasso regression clump down the co-efficient to be irrelevant and provide you with a sparse model which is not the case with Ridge regression.

As one keeps on increasing the alpha values it imposes more and more penalty on the co-efficient and hence variable selection is performed by Lasso

## Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

**Solution:**

- The generalized regression process includes the following steps:
  - Conduct exploratory data analysis by examining scatter plots of explanatory and dependent variables.
  - Choose an appropriate set of functions that seem to fit the plot well and build models using them.

- While constructing the regression model, instead of using the raw explanatory variables in the current form, we create some function of the explanatory variables to best explain the data points.
- Raw attributes, as they appear in the training data, may not be best suited to predict the response variable value. This way we can build a robust & generalized model
- in order to best fit a non-linear curve to the data, we figure out how close is the set of predicted values to the set of given values for the response variable, just like in simple linear regression. The only additional thing you need to do is to pick the right features or functions of raw attributes and run a regression on features instead running it on raw attributes.
- Model needs to be made more generalizable and robust so that they are note affected by the outliers in the training data and the test accuracy is not less than the training score
- A model should not only be accurate to the given data set but should also perform well on the unseen(test) data
- Outliers that doesn't make sense should be removed this will increase the accuracy of the prediction made by the model