

# Subjective Questions

**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 double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?**

I found below analysis's result based on jupyter notebook

## **Ridge Regression:**

Optimal Value of lamda: 3  
Ridge regression train r2: 0.89

Ridge\_Train\_MSE : 0.016

Ridge regression test r2: 0.87

Ridge\_Test\_MSE : 0.021

## **Top 5 most significant variables are:**

Neighborhood\_Timber  
Neighborhood\_StoneBr  
Neighborhood\_Somerst  
Neighborhood\_SawyerW  
Neighborhood\_Sawyer

## **Lasso Regression:**

Optimal Value of lamda: 0.0004

Lasso Regression train r2: 0.8934

Lasso\_Train\_MSE : 0.016

Lasso Regression test r2: 0.8718

Lasso\_Test\_MSE : 0.021

## **Top 5 most significant variables are:**

- ('Neighborhood\_Timber', 0.145)
- ('Neighborhood\_StoneBr', 0.101),
- ('Neighborhood\_Somerst', 0.1),
- ('Neighborhood\_SawyerW', 0.095),

- ('Neighborhood\_Sawyer', 0.064),

If optimal value is doubled, then

### **Ridge Regression:**

Optimal Value of lamda: 6

Ridge regression train r2: 0.8934

Ridge regression test r2: 0.8706

Ridge\_Train\_MSE : 0.016

Ridge\_Test\_MSE : 0.021

#### **Top 5 most significant variables are:**

- ('Neighborhood\_Timber')
- ('Neighborhood\_StoneBr')
- ('Neighborhood\_Somerst')
- ('Neighborhood\_SawyerW')
- ('Neighborhood\_Sawyer')

### **Lasso Regression:**

Optimal Value of lamda: 0.0008

Lasso Regression train r2: 0.8923

Lasso Regression test r2: 0.8726

Lasso\_Train\_MSE: 0.0169

Lasso\_Test\_MSE : 0.0209

#### **Top 5 most significant variables are:**

- ('Neighborhood\_Timber', 0.119)
- ('Neighborhood\_StoneBr', 0.102)
- ('Neighborhood\_Somerst', 0.093)
- ('Neighborhood\_SawyerW', 0.077)
- ('Neighborhood\_Sawyer', 0.054)

**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?**

The values calculated for both Ridge and Lasso Regression are almost identical. However, I would prefer to choose Lasso regression as it does Feature selection.

**Question 3: After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?**

Dropping the first 5 predictors and apply the regression with rest of the data.

**Top 5 most significant variables are:**

- ('Neighborhood\_SWISU', 0.054)
- ('Neighborhood\_OldTown', 0.043)
- ('Neighborhood\_NridgHt', 0.042)
- ('Neighborhood\_NoRidge', 0.04)
- ('Neighborhood\_NWAmes', 0.039)

**Question 4: How can you make sure that a model is robust and generalized? What are the implications of the same for the accuracy of the model and why?**

A model can be made robust and generalized by making sure that the data is not being impacted by outliers. Also, we need to model such that the testing accuracy is not less than the training set. Regularization helps to build a stable model which we have seen in this assignment.