

**Question 1:**

**Problem Statement - Part II**

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?

**Answer1:**

Optimal value of alpha for ridge: **1**

Optimal value of alpha for ridge: **0.0001**

After make the double alpha for ridge and lasso i.e. **2 and 0.0002**

**For Ridge:**

**Before Doubling:**

Ridge regression train r2: **0.9339**

Ridge regression test r2: **0.9096**

**After Doubling:**

Ridge regression train r2: **0.9315**

Ridge regression test r2: **0.9069**

Top features: Id, LotArea, BsmtUnfSF, BsmtFinSF1, OverallQual

**For Lasso:**

**Before Doubling:**

Lasso Regression train r2: **0.9345**

Lasso Regression test r2: **0.9131**

Top features: LotArea, KitchenAbvGr, OverallCond, MSZoning\_RH, MSZoning\_RM

**After Doubling:**

Lasso Regression train r2: **0.9311**

Lasso Regression test r2: **0.9138**

Top features: LotArea, KitchenAbvGr, OverallCond, YearBuilt, HeatingQC

**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?

**Answer 2:**

We will choose **Lasso** as its giving **feature selection** option also. It has removed unwanted features from model without affecting the model accuracy. Which makes are model generalized and simple and accurate. Additionally, the r2\_score of lasso is slightly higher than lasso for the test dataset so we will choose lasso regression to solve this problem

### **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?

### **Answer 3:**

**Previous top 5 features were:** LotArea, KitchenAbvGr, OverallCond, YearBuilt, HeatingQC

### **R2 after dropping:**

**Lasso Regression train r2: 0.9215**

**Lasso Regression test r2: 0.8951**

**Next top 5 features** after dropping 5 main predictors are: OverallQual, KitchenQual, YearBuilt, HeatingQC, RoofMatl\_Metal

### **Question 4:**

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

### **Answer 4:**

The model should be generalized so that the test accuracy is not lesser than the training score. The model should be accurate for datasets other than the ones which were used during training. Too much importance should not be given to the outliers so that the accuracy predicted by the model is high. To ensure that this is not the case, the outliers analysis needs to be done and only those which are relevant to the dataset need to be retained. Those outliers which it does not make sense to keep must be removed from the dataset. If the model is not robust, it cannot be trusted for predictive analysis.