**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?**

**Answer :**

Optimal value of alpha for ridge is : 1.0

Optimal value of alpha for Lasso is : 0.001

If we use double the value of alpha for both Ridge and Lasso then the coefficients will become more simpler and the regularization will increase, due to which the bias will increase and variance will decrease, we saw in our case as well that the training accuracy is going down due to the same which is the indication of bias getting increased. In other words model is basically becoming more simpler.

The most important predictor variables after the change is implemented would be:

a) GrLivArea: Above grade (ground) living area square feet

b) BsmtFullBath : Basement Full Bathrooms

c) BsmtQual: Evaluates the height of the basement

d) Neighborhood\_IDOTRR : Physical locations within Ames city limits - Iowa DOT and Rail Road

e) LowQualFinSF : Low quality finished square feet (all floors)

**------------------------------------------------------------------------------------------------------------------------------------------**

**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 :**

We will use Lasso because in Lasso many of the redundant features will be dropped as the coefficients corresponding to the redundant features will essentially become zero. In other words it would be easy for feature selection technique when we use Lasso.

**------------------------------------------------------------------------------------------------------------------------------------------**

**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 :**

The new five most important variables now would be :

a) LotShape : General shape of property

b) Fireplaces : Number of fireplaces

c) BsmtHalfBath : Basement half bathrooms

d) 1stFlrSF : First Floor square feet

e) BsmtExposure : Refers to walkout or garden level walls

**------------------------------------------------------------------------------------------------------------------------------------------**

**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 :**

For making sure that the model is robust and generalisable we will check the bias and the variance of the model. If there isn't a large gap between the test(unseen data) and train accuracy then we can say that the model we have build is generalisable. We would treat the outliers and the multicollinearity as well for making the model more robust. If model is not that generalisable then the bias would be very low but the variance will be very high for unseen data which is a clear indication of the overfitting.

The model accuracy would be differing a lot on the unseen data when compared with the train data that is it will be very low on the unseen data if the model is not robust and generalisable.

**------------------------------------------------------------------------------------------------------------------------------------------**