# Advanced Regression Objective 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?**

The optimal value of alpha for Lasso and Ridge are 0.001 and 10.0 respectively. For these optimal values the R^2 score is 91.6% and 91.8% for both Lasso and Ridge respectively.

The most important 10 predictor variables for above alpha values are: -

```
`Neighborhood_Crawfor`
`Neighborhood_StoneBr`
`GrLivArea`
`Neighborhood_NridgHt`
`OverallQual`
`Exterior1st_BrkFace`
`Functional_Typ`
`SaleCondition_Normal`
`OverallCond`
`SaleType_New`
```

When I double the value of alphas I observed that R^2 score is very slightly increased for Lasso and is 91.8%. Whereas, almost same in R^2 for Ridge which is 91.79%.

The most important 10 predictor variables for double the alpha values are: -

```
GrLivArea
Neighborhood_Crawfor
Neighborhood_StoneBr
OverallQual
OverallCond
Neighborhood_NridgHt
Functional_Typ
SaleType_New
```

```
SaleCondition_Normal
Condition1_Norm
```

**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 main advantage of ridge regression is coefficient shrinkage and reducing model complexity. However, along with shrinking coefficients, lasso performs feature selection as well as some of the coefficients become exactly zero.

Also, since Ridge includes all the features, it is not very useful in case of extremely high number of features as it will pose computational challenges.

Since Lasso provides sparse solutions, it is generally the model of choice for modelling cases where the number of features is extremely huge. In such a case, getting a sparse solution is of great computational advantage as the features with zero coefficients can simply be ignored.

*Considering above advantages for Lasso, I will choose Lasso Regression.*

**Question 3: -**

**After building the model, you realized 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?**

After excluding the top 5 predictor variables from the previous Lasso model, below are the top 5 predictor variables in the new Lasso model.

```
`2ndFlrSF`
`1stFlrSF`
`Functional_Typ`
`MSZoning_FV`
`MSZoning_FV`
```

**Note: - New model code is also present in the Python Notebook.**

**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?**

Robustness is the property that tested on a training sample and on a similar testing sample, the performance is close.

To get our model robust and generalizable, we have to choose a simple and flexible model. That means we should not go for high degree polynomial. High degree polynomial models are more complex models and more accurate on training data which leads to overfitting.

A flexible one will learn from data a lot. So usually avoiding accurate model and preferring a robust one is better.