

# Advance Regression - 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?

The Optimal value of alpha for Ridge Regression is 50 and for Lasso Regression is 0.0001.

Generally, when we increase the alpha value, it will decrease the coefficients value and R2 square value gradually and as we keep decreasing this value model will under fit eventually.

When we increased the alpha value to twice the optimal value then there would be more penalization and decrease in the coefficients. As we saw in our model, lets take OverallQual and Overall Cond. Here we saw that the RidgeModel coefficient decreased from 0.099552 to 0.91808 and for LassoModel it decreased from 0.110363 to 0.110424

<b>OverallQual</b>	0.110290	0.099552	0.110363	0.110363	0.099552	0.110290	0.091808	0.110424
<b>OverallCond</b>	0.046864	0.044664	0.046767	0.046767	0.044664	0.046864	0.042554	0.046671

Although the coefficients value decreased. the Overall Quality is still the variable with most importance

## 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?

Both Ridge and Lasso Model provided similar results with out much variance. However, Lasso Regression would be a better model over Ridge as Ridge regression shrinks the coefficients towards zero, but it will not set any of them exactly to zero. However, Lasso Regression would set co-efficient to zero this would help in feature elimination and help build more robust model.

### 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?

The 5 most important predictor variables for the original model are

1. Overall Quality
2. Total rooms above grade (does not include bathrooms)
3. Overall condition of the house
4. Full bathrooms above grade
5. Size of garage in car capacity

The 5 most Important predictor after the above variables are excluded are:

1. Kitchen quality
2. GarageArea: Size of garage in square feet
3. BedroomAbvGr: Bedrooms above grade (does NOT include basement bedrooms)
4. BsmtQual: Evaluates the height of the basement
5. FireplaceQu: Fireplace quality

### 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?

If the model is not robust, it cannot be trusted for predictive analysis. A model has to be made robust and generalizable so that the outliers in the training data do not impact it. When the accuracy of a model when dealing with real time data, if not generalized, can fall drastically to below 20% or so even when the accuracy with training and test would be over 80%. To ensure generalization, the bias and variance must be kept low. It must also be generalizable so that the training score is lesser than the test accuracy. The model should be accurate for unseen data as well. Too much weightage should not give to the outliers so that the accuracy predicted by the model is high. To ensure that this is not the case, the outlier 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. This would help increase the accuracy of the predictions made by the model. Confidence intervals can be used i.e. typically 3-5 standard deviations. This would help standardize the predictions made by the model.