

# Advanced 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?

### Answer:

In my model, the optimal value of for Ridge regression is 10 and for Lasso regression is 0.01.

If we double the value of alpha, for Ridge regression the magnitude of coefficients will decrease further towards 0, but it will not become zero, whereas in the Lasso regression the coefficients will become zero for a few more parameters.

After the change, the most important predictor variables for Ridge regression are: Neighborhood\_Stone Br, GarageArea, Neighborhood\_NridgHt, TotalBsmtSF, GrLivArea, KitchenQual, Neighborhood\_Names, Neighborhood\_Edwards, BldgType\_TwnhsE, GarageFinish, and for Lasso regression are: TotalBsmtSF, SaleType\_New, MSZoning\_RM, GarageType\_Attchd, GrLivArea, Neighborhood\_NAMES, Neighborhood\_OldTown, KitchenQual, SaleCondition\_Partial, RoofStyle\_Gable.

## 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:

In my model, I observed the below R2 scores.

#### Ridge Regression:

Train set: 92% Test set: 88%

#### Lasso Regression:

Train set: 86% Test set: 85%

Even though the Ridge model performs better than the Lasso in terms of R2 score, I will still choose Lasso model because it helps with feature elimination significantly. Since our dataset had more than 200 columns (post dummy variable generation), feature elimination is an important aspect to keep the model complexity to simple and easy interpretations. Hence, I will choose Lasso model over Ridge.

### 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:

In my Lasso model, the five most predictor variables are the below.

OverallQual, GrLiveArea, OverallCond, GarageArea, BsmIFullBath

If we create another model excluding the above variables, then the five most predictor variables are the following:

Fireplaces, FullBath, TotalBsmISF, LotArea, MSZoning\_RL

### 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:

To make a model robust, we need to make sure it is not complex and keep as minimum predictor variables as possible. To make the model generalisable, we need to make sure it is not overfitting on the train data. Regularization techniques will help in both these scenarios.

The accuracy of the model will be high, if we overfit the model, but then to generalize it well, we need to make sure it is not overfitting, which might lead to lesser accuracy of the model, but accuracy will match better on both train and test sets.