

## Assignment Part-II

### 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:

- i) Following are optimal value of alpha for ridge and lasso regression:

	Metric	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.858983	0.861861
1	R2 Score (Test)	0.848937	0.851301
2	RSS (Train)	21.964878	21.516597
3	RSS (Test)	11.632177	11.450142
4	MSE (Train)	0.021513	0.021074
5	MSE (Test)	0.026557	0.026142

Optimal Value of alpha for ridge regression is 50.0 and for Lasso is 0.001. Top 5 most important predictors are as below.

### For Ridge

	Features	Values
0	constant	13.428
12	OverallQual	0.086
38	GarageCars	0.077
28	BsmtFullBath	0.070
34	Fireplaces	0.060
32	TotRmsAbvGrd	0.059

### For Lasso

	Features	Values
0	constant	13.646
12	OverallQual	0.087
38	GarageCars	0.086
28	BsmtFullBath	0.083
29	FullBath	0.065
34	Fireplaces	0.064

- ii) After we double the value of Alpha (i.e. ridge regression is 100.0 and for Lasso is 0.002), following changes are observed:

If we choose to double the value of alpha for ridge regression, we observe that the coefficients value decreases and for lasso some more coefficients turns to 0 which are less important.

1. Little change in R2 values for Both Ridge and Lasso (Train, Test Data)

	Metric	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.855379	0.857019
1	R2 Score (Test)	0.845317	0.850990
2	RSS (Train)	22.526247	22.270833
3	RSS (Test)	11.910918	11.474094
4	MSE (Train)	0.022063	0.021813
5	MSE (Test)	0.027194	0.026197

2. One variable changes in Top 5 Lasso Top 5 Predictors

iii) Following are Top 5 most important predictors after the change:

### For Ridge

	Features	Values
0	constant	12.368
12	OverallQual	0.085
38	GarageCars	0.068
32	TotRmsAbvGrd	0.060
28	BsmtFullBath	0.059
34	Fireplaces	0.056

### For Lasso

	Features	Values
0	constant	11.625
12	OverallQual	0.090
38	GarageCars	0.081
28	BsmtFullBath	0.079
34	Fireplaces	0.065
32	TotRmsAbvGrd	0.055

### 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:**

Both ridge and Lasso model train and test accuracy score is quite similar so we choose Lasso over Ridge because with Lasso regression we can do feature selection as well.

### 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.

What are the five most important predictor variables now?

**Answer:**

After dropping five of the most important predictors and again retraining the model we observed the below five most important predictor Variables.

### For Ridge

	Features	Values
0	constant	25.002
27	FullBath	0.207
28	HalfBath	0.129
26	CentralAir	0.090
36	PavedDrive	0.047
12	OverallCond	0.045

### For Lasso

	Features	Values
0	constant	24.316
27	FullBath	0.229
28	HalfBath	0.140
26	CentralAir	0.107
36	PavedDrive	0.045
12	OverallCond	0.044

#### 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:

A Model is robust when any variation in data doesn't affect its performance. A model is generalisable when it performs similar with new unseen data. Model should be as simple as possible. Means Bias and Variance should be as low as possible. To make sure a model is robust and generalizable, we have to take care it doesn't overfit and underfit. This is because an overfitting model has very high variance and a smallest change in data affects the model prediction heavily. Such a model will identify all the patterns of a training data, but will fail to pick up the patterns in unseen test data. We look at it from the perspective of Accuracy, a too complex model will have a very high accuracy. So, to make our model more robust and generalizable, we will have to decrease variance which will lead to some bias. Addition of bias means that accuracy will decrease. We have to find some balance between Bias and Variance and this can be achieved by Regularization Techniques like Ridge and Lasso.