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:

The optimal value of alpha for ridge and lasso regression are

Regression	Optimal Value of Alpha
Ridge	3
Lasso	100

When I double the value of alpha for both ridge and lasso regression. Below are the readings:

	Metric	Ridge Regression(3)	Ridge Regression(6)	Lasso Regression(100)	Lasso Regression(200)
0	R2 Score (Train)	9.281289e-01	9.278579e-01	9.266148e-01	9.241128e-01
1	R2 Score (Test)	8.788714e-01	8.793047e-01	8.803521e-01	8.827945e-01
2	RSS (Train)	3.156737e+11	3.168637e+11	3.223239e+11	3.333132e+11
3	RSS (Test)	3.147568e+11	3.136307e+11	3.109091e+11	3.045624e+11
4	MSE (Train)	1.853367e+04	1.856857e+04	1.872788e+04	1.904445e+04
5	MSE (Test)	2.826438e+04	2.821377e+04	2.809109e+04	2.780290e+04

Most important predictors after change for ridge regression

Ridge (α=3)		Ridge (α=6)	
GrLivArea	23434.360603	GrLivArea	22751.803929
OverallQual	10947.925302	OverallQual	11085.181237
RoofStyle_Gable	10113.624420	TotalBsmtSF	10178.006566
TotalBsmtSF	9983.685473	RoofStyle_Gable	8233.434554
RoofStyle_Hip	9898.648574	RoofStyle_Hip	8125.825595
LotArea	7434.218782	LotArea	7344.988551
MSZoning_RL	6908.805340	Neighborhood_NridgHt	5348.830239
Neighborhood_NridgHt	5244.365894	BsmtFullBath	5001.176775
MSZoning_FV	5227.899762	BsmtExposure_Gd	4903.818301
GarageType_Detchd	5009.740908	MSZoning_RL	4823.909116

Most important predictors after change for lasso regression

Lasso (α=100)		Lasso (α=200)	
GrLivArea	23442.632217	GrLivArea	23026.851354
OverallQual	11572.474756	OverallQual	12184.867158
TotalBsmtSF	10314.535511	TotalBsmtSF	10833.846186
LotArea	6937.785882	LotArea	6653.491310
Neighborhood_NridgHt	6039.890296	Neighborhood_NridgHt	6482.650765
BsmtFullBath	4901.424103	Neighborhood_NoRidge	4904.608537
Neighborhood_NoRidge	4886.802620	BsmtFullBath	4864.987992
BsmtExposure_Gd	4799.139332	BsmtExposure_Gd	4640.396118
Neighborhood_StoneBr	4480.233061	Neighborhood_StoneBr	4640.143227
Neighborhood_Crawfor	3890.635784	Neighborhood_Crawfor	3735.044390

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:

The feature selected for Ridge and Lasso regressions are

Ridge(α=3/6)	Lasso(α=100)	Lasso(α=200)
124	108	99

I will be selecting Lasso Regression for alpha value of 200.

1. It has minimal (99) variables for interpretation. Since Lasso shrinks some of the variable coefficients to 0 thus performing variable selection
2. R2Score, RSS and MSE are better for lasso.

	Metric	Ridge Regression(3)	Ridge Regression(6)	Lasso Regression(100)	Lasso Regression(200)
0	R2 Score (Train)	9.281289e-01	9.278579e-01	9.266148e-01	9.241128e-01
1	R2 Score (Test)	8.788714e-01	8.793047e-01	8.803521e-01	8.827945e-01
2	RSS (Train)	3.156737e+11	3.168637e+11	3.223239e+11	3.333132e+11
3	RSS (Test)	3.147568e+11	3.136307e+11	3.109091e+11	3.045624e+11
4	MSE (Train)	1.853367e+04	1.856857e+04	1.872788e+04	1.904445e+04
5	MSE (Test)	2.826438e+04	2.821377e+04	2.809109e+04	2.780290e+04

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 five most important predictor variables for Lasso Regression (α=100 – optimal value) are

1. TotRmsAbvGrd: Total rooms above grade
2. GarageArea: garage area
3. FullBath: Full bathrooms above grade
4. Fireplaces: Number of fire places
5. Neighborhood_NoRidge: NorthRidge

```
print(betas['Lasso(100)'].sort_values(ascending=False).head())
```

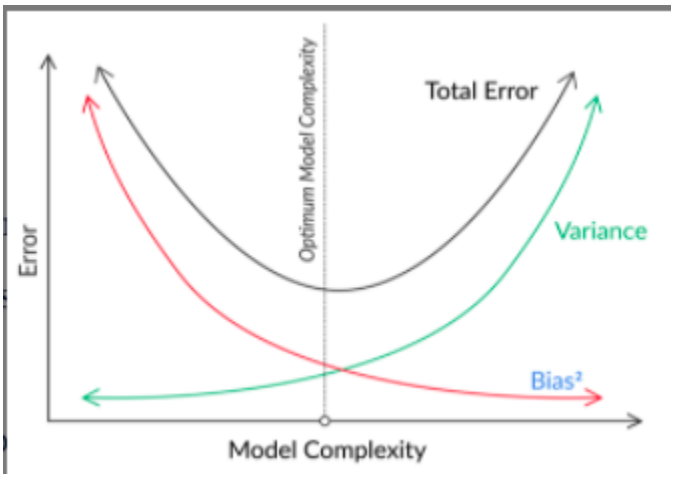
```
TotRmsAbvGrd            16416.340816
GarageArea               7866.902716
FullBath                 7690.598467
Fireplaces               7581.091049
Neighborhood_NoRidge     7271.645043
```

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:

We can say that a model is robust and generalized when both Bias and Variance is low. In this case the model performs well in both test and trained data.



When Bais is low and variance is high that means model is too complex and may leads to overfitting. This kind of model performs well on train data but not on test data.

When Bais is high and Variance is low that means model is too simple and may leads to underfitting. This kind of model neither perform well on train data nor perform well on test data.

Regularization helps with managing the model complexity by essentially shrinking the model coefficients estimates towards 0. This discourage the model being too complex and thus avoiding the risk of overfitting.