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:

Optimal value of alpha for ridge is 2.0
Optimal value of alpha for lasso is 0.0001

After doubling the value of alpha for both ridge and lasso, following are the r2 score, rss and mse values. There's not a big difference in metrics after doubling the alpha value. It could be because of having low value for alpha.

Metric	Ride Regression (alpha – 0.4)	Lasso Regression (alpha - 0.0002)
R2 score (train)	0.870853	0.866943
R2 score (test)	0.840605	0.843009
RSS (train)	86.321091	86.321091
RSS (test)	40.588000	40.588000
MSE (train)	0.032677	0.032677
MSE (test)	0.034244	0.034244

The most important predictor variables after the change is implemented,
Ridge Regression:

#	Feature	Coefficient
1	OverallQual	0.179342
2	TotRmsAbvGrd	0.128423
3	GarageCars	0.092059
4	Neighborhood_StoneBr	0.078471
5	Neighborhood_NoRidge	0.076238
6	GarageArea	0.072821
7	FullBath	0.069749
8	Neighborhood_NridgHt	0.061532
9	BsmtQual_Ex	0.054206

Lasso Regression:

#	Feature	Coefficient
1	OverallQual	0.239815

2	TotRmsAbvGrd	0.148355
3	GarageCars	0.094765
4	Neighborhood_StoneBr	0.081813
5	Neighborhood_NoRidge	0.079124
6	Neighborhood_NridgHt	0.063307
7	GarageArea	0.059730
8	FullBath	0.058502
9	BsmtFullBath	0.055096

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:

Optimal value of lambda for ridge is 2.0
Optimal value of lambda for lasso is 0.0001

R2 score and Mean squared value has negligible difference between ridge and lasso regression.
Ridge regression R2 score (for test data) – 0.840605
Lasso regression R2 score (for test data) – 0.843009

Ridge regression MSE (for test data) – 0.034244
Lasso regression MSE (for test data) – 0.034244

I would choose **lasso regression** here since there's no big difference in r2 score or MSE and lasso reduces some of the features (coefficient as 0)

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 in the lasso model:
1. OverallQual
2. TotRmsAbvGrd
3. GarageCars
4. Neighborhood_StoneBr
5. LotArea

After excluding the above five most important predictor variables, created another lasso model. Now, the five most important predictor variables are,
1. GarageArea
2. FullBath
3. Neighborhood_NoRidge
4. RoofStyle_Shed
5. BedroomAbvGr

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:

As per Occam's Razor rule,
- Simpler the model, more generalizable it is.
- Simpler models are more robust.
- Making model as simple as possible but no simpler.

To create simpler models, need to balance variance and bias. We can also use regularization which prevents model from becoming complex
- Variance – How sensitive is the model to input data
- Bias – How much error the model likely to make in test data

Bias-Variance Tradeoff