

Subjective Questions (Submission by Satya Prakash)

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 or Ridge is 0 and Lasso is 0.0009
- After doubling the alpha values, the model performance almost remains same.
- The most important predictor variables in Ridge after the change are:

Ridge Doubled Alpha Co-Efficient	
OverallQual	0.367308
LotArea	0.195262
TotRmsAbvGrd	0.115728
LotFrontage	0.091084
GarageYrBlt	0.079909
OverallCond	0.079213
GarageCars	0.075778
Fireplaces	0.067785
GarageArea	0.059203
MasVnrArea	0.047154
BedroomAbvGr	-0.001983
KitchenAbvGr	-0.069855

- The most important predictor variables in Lasso after the change are:

Lasso Doubled Alpha Co-Efficient	
OverallQual	0.368065
GarageCars	0.116199
TotRmsAbvGrd	0.076931
Fireplaces	0.072190
GarageYrBlt	0.036194
GarageArea	0.024713
LotFrontage	0.000000
LotArea	0.000000
OverallCond	0.000000
MasVnrArea	0.000000
BedroomAbvGr	0.000000
KitchenAbvGr	-0.000000

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:

For me both lasso and ridge has given similar performance so it is a difficult choice based on statistics. But considering that we have too many features, I would prefer lasso. It will help with feature elimination and make a simple and robust model.

	Metric	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.829734	0.829482
1	R2 Score (Test)	0.783199	0.784552
2	RSS (Train)	2.451437	26563.266489
3	RSS (Test)	1.386348	5058.950675
4	MSE (Train)	0.049096	0.049133
5	MSE (Test)	0.056324	0.056148

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:

After removing the top 5 features ('OverallQual', 'LotArea', 'TotRmsAbvGrd', 'LotFrontage', 'GarageYrBlt') and building the lasso model below are the five most important predictor variables.

Lasso Co-Efficient	
GarageCars	0.240194
Fireplaces	0.148310
GarageArea	0.141666
MasVnrArea	0.088077
BedroomAbvGr	0.068846

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:

For the model to be robust and generalizable, it should be as simple as possible. It should be able to identify the patterns in the data than memorizing them and avoid overfitting. This can be achieved by using regularization method. Robustness can be tested by multiple validation methods on the unseen data. Simpler the model the more the bias but less variance and more generalizable. Use Bias-Variance trade-off. Its implication in terms of accuracy is that a robust and generalizable model will perform equally well on both training and test dataset.