

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?

Optimal value of alpha for Ridge regression: 500

Optimal value of alpha for Lasso regression: 500

Therefore both regressions yield similar alpha value.

Effect of doubling alpha value (i.e. changing Alpha from 500 to 1000):

Below are R2 values for different Alpha settings:

Alpha = 500

Metric	Ridge Regression	Lasso Regression
R2 Score (Train)	0.8834	0.9233
R2 Score (Test)	0.8552	0.8443

Alpha = 1000

Metric	Ridge Regression	Lasso Regression
R2 Score (Train)	0.8585	0.8919
R2 Score (Test)	0.8408	0.8418

R2 value for test dataset is bit better with Alpha =500 in both Ridge and Lasso regression cases.

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?

Observation:

- R2 scores of train data and test data are closer in case of Ridge regression

Ridge regression should yield better results in this case.

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?

Based on coefficient values, followings are most important variables:

- 1) OverallQual
- 2) GrLivArea
- 3) Neighborhood_NoRidge
- 4) 1stFlrSF
- 5) Neighborhood_NridgHt

Next important variables will be:

- 1) TotRmsAbvGrd
- 2) GarageCars
- 3) TotalBsmtSF
- 4) 2ndFlrSF
- 5) Fireplaces

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?

We should keep model simple to make it robust and generalisable. Over complex model tend to overfit and will start showing poor results on test data.