

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?

Ans :

Optimal value of alpha :

Ridge regression = 50

Lasso regression = 0.001

If the alpha is doubled then :

For ridge regression – Not much change in R2 score of Train and test data is observed.

For lasso regression – R2 score almost matching for train and test data

Regression	Alpha	R2 score	
		Train	Test
Ridge	50	0.9038	0.8718
	100	0.8961	0.8693
Lasso	0.001	0.8433	0.8431
	0.002	0.8611	0.8503

Most important predictor :

For Ridge regression – OverallQual

For Lasso Regression – OverallQual

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?

Ans :

Though the model performance by Ridge Regression was better in terms of R2 values of Train and Test, it is better to use Lasso, since it brings and assigns a zero value to insignificant features, enabling us to choose the predictive variables. It is always advisable to use simple yet robust model.

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?

Ans :

Five most important predictor variables are as follows :

- 1) OverallQual
- 2) GrLivArea
- 3) PropAge
- 4) OverallCond
- 5) GarageArea

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?

Ans :

Model should be as simple as possible, though its accuracy will decrease but it will be more robust and generalizable. It can be understood by bias-variance tradeoff. The simpler the model, more will bias but less variance and more generalizable will be the model. It will perform equally on both training and test data set

It is important to have balance between bias and variance to avoid overfitting and underfitting.