

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:

Ridge Alpha: 7

Lasso Alpha: 0.0001

- When we double the value of alpha for Ridge Regression, the model will penalise more as a result of which model will be more generalised. There are chances of more error in train and test data.
- When we double the value for Lasso Regression, it will penalise model and decreasing the R2 value.
- Predictor variables after the changes are:
 1. LotFrontage
 2. Street
 3. Neighbourhood
 4. FullBath
 5. MSZoning

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:

We will compare R2 value for both Ridge and Lasso regression

	Ridge regression	Lasso Regression
R2 score - Train	0.958	0.942
R2 score - Test	0.928	0.924

You can see the R2 score on test data is only marginally better on Ridge regression compared to lasso. We will choose lasso regression as it will help in feature elimination and build a 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?

Answer:

The five most important variables now are:

1. LotArea
2. OverallQual
3. LandContour
4. GarageYrBlt
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?

Answer:

The model should be simple, its accuracy however will decrease but it will be robust and generalisable. On a simple and generalised model, The accuracy on test data is not much lower compared to training data set.

It is important to have balance between Bias and Variance to overcome overfitting and underfitting.