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 1

Optimal value – Ridge 10 and Lasso – 0.0004. Doubling alpha (Lambda) will cause greater penalty to be enforced and although the variance will decrease but the bias will rise. This will cause to model to underfit the data.

Important variables after the change in implemented:

1. MSZoning_RL
2. GrLivArea
3. MSZoning_RM
4. OverallQual
5. MSZoning_FV

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 2

Lasso will be applied as it gives lower MSE and also nocks out some of the variables making computation and understanding easier

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 3

1. GrLivArea
2. OverallQual
3. OverallCond
4. TotalBsmtSF
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?

Answer 4

We need to find a balance b/w bias and variance and use the cut-off for the final model