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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer:

Optimal value of alpha for Ridge is 10 and for lasso is 100.

After doubling the alpha values following changes are observed:

1. MSE/RMSE for both lasso and ridge decreases.
2. R2 increases

Most important predictor for both remains the same i.e., *GrLivArea*.

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:

Lasso. Because model evaluation reveals that lasso is slightly better and since it also does feature elimination results in simpler mode.

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:

Five most important predictors before are Overall Quality of Excellent and Very Excellent, Good Pool Quality, Above ground living area, Proximity to off site feature like Park, Greenbelt, etc.

After removing this feature and building new models most important features are - Neighbourhood of Northridge, severely damaged home functionality, Metal Roof Material, First floor and Second floor area in sq. ft.

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:

Residual analysis can be done to make sure that assumptions of regression are met. A good model should be able to balance the bias-variance trade-off for it to be able to generalise and provide accurate prediction on unseen data.