Assignment-based Subjective Questions

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?

The optimal value for alpha obtained is 20 for ridge regression and 100 for lasso regression.

There is a slight decrease in the r squared value on doubling the alpha value.

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?

The R squared value for Lasso is slightly better than Ridge regression. Also since the number of features are a lot, feature selection done by Lasso is also an advantage.

Lasso regression in this case would be a better option.

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?

The five most important features after removing the initial 5 imortatnt features are GrLivArea, Exterior2nd_ImStucc, Exterior1st_BrkFace, LandContour_HLS and LandContour_Lvl.

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?

For making the model robust and generalisable it is imp to avoid overfitting where the model essentially memorizes the dataset.

Regularization helps in keeping the model robust and more general.

A robust and general model has low variance which is achieved by making the model less complex by reducing the number of features.

This will reduce the accuracy as we are removing the features , but will make the model more general.