Housing Price Prediction_Advance Regression Assignment

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:

The optimal value of alpha for the Ridge regression was found to be '10' and for the Lasso regression was found to be '0.01'.

If we double the value of alpha for the Ridge regression to be '20' and for the Lasso regression to be '0.02', then following is observed:

1. The R_squared for value Ridge regression had a minimal decrease from 0.773 to 0.755 however the R_squared for value Lasso regression had a considerable decrease from 0.738 to 0.671.
2. The model coefficients value for both Ridge regression and Lasso regression had a considerable decrease for all the features.

The most important predictor for Ridge Regression is 'The overall condition of the house being 'Very Excellent'' that is same as the previous case.

The most important predictor for Lasso Regression previously was 'The overall condition of the house being 'Very Excellent'', however after changing the alpha value, the new most important predictor is the 'garage having 3 car capacity'.

_____

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:

The optimal value of alpha for the Ridge regression was found to be '10' and for the Lasso regression was found to be '0.01'

The Ridge regression should be preferred to be applied for the prediction of house price since it Ridge regression has a higher R_squared compared to the Lasso and Ridge also contains more features in the model for prediction of the target variable & allows better understanding of the house price features.

_____

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:

The five most important predictors currently found from the existing model are:

- The overall condition of the house being 'Very Excellent', 'Excellent' & 'Very Good' contributes majorly for the house price.
- The 'garage having 3 car capacity' also contributes majorly for the house price.
- '3 Full Bathrooms' above Grade contributes significantly for the house price.
- The houses at the location 'North ridge' also contributes significantly for the house price.
- The houses with high 'First Floor square feet' also contributes significantly for the house price.

If the above predictors are found absent, then the next most important predictors are as follows:

- The houses with high 'First Floor square feet'.
- The 'Basement with Good Exposure' to walkout or garden level walls.
- The overall condition of the house being 'Good'.
- The houses at the location 'Crawford'.
- The newly constructed houses.

_____

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:

The model can be ensured as robust and generalisable:

1. If the difference between the accuracy of the training data and the test data is not considerably high.
2. If the model coefficients are relatively low.
3. If the bias and variance is balanced, both of them should not be relatively high.

A robust and generalised model compromises a little amount of bias to get a significant decrease in the variance. The overall accuracy of such models is comparatively low however the prediction accuracy for unseen data is much better.

_____