SURPRISE HOUSING PRICE ASSIGNMENT SUBJECTIVE QUESTION SOLUTION

Q1. 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?

Ans 1.

1. The optimal value of Lambda for –

    Ridge Regression = 10

    Lasso Regression = 0.001

2. When we doubled the value of alpha both both ridge and lasso Regression i.e., 20 & 0.002 respectively.

    We will see following changes:-

- The value of R2 Score (Train) decreases for both Ridge and Lasso
- The value of R2 Score (Test) decreases for Lasso and remains same for Ridge.
- After the changed are done, the most important predicting variables will be – GrLivArea, OverallQual_8, OverallQual_9, same as that in previous case, just coeff changes.

Q2. 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?

Ans 2.

It depends on the situation of case like –

1. If we have too many variables and one of our primary goals is to select features, then we will use Lasso
2. If we want to get reduction of coefficient magnitude, then we will use Ridge Regression.

In our case, we have done both the methods and compared the model parameters which are almost having same values for Train and Test data both. So, we will use Lasso Regression, to get most important features to be selected in predicting housing price.

Q3. 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?

Ans 3.

After removing our first five predictors in Lasso, we get following top 5 new predictors – 2ndFlrSF, Functional_Typ, 1stFlrSF, MSSubClass_70, Neighbourhood_Somerst

Q4. 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?

Ans 4.

A Robust model is one which does not affect its performance much with any changes in data and a Generalizable model is that which easily adapts to new one without seeing previous data with same distribution which used to create model.

To make our model Robust and Generalizable, we should have to take care of following parameters:-

a.   Overfit – it should not be overfit, i.e., it should have low variance because it will fail to perform on test data if variance is too high

b.   Complexity – it should not be too complex with low variance and bias

c.   Accuracy – if model is too complex, accuracy will be high. We need to decrease variance & add bias which decreases accuracy.

d.   Last, we have to take care of balance between both accuracy and complexity, i.e., both should neither be too low nor be too high.