House Price Prediction – Advanced Regression Assignment Part 2

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 for alpha is 2 for Ridge and for lasso is 0.002. The change in the model as follows :

Ridge regression : R2score on training data has decreased but it has increased on testing data

Lasso regression : R2score of training data has decrease and it has increase on testing data

The most important predictor variable after change :

- Lot Area
- OverallQual
- OverallCond
- YearBuilt
- BsmtFinSF1
- TotalBsmtSF

House Price Prediction – Advanced Regression Assignment Part 2

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

In the assignment , I have used build three models with different regression technique linear , ridge and lasso and according to r2_score of all the model Ridge Regression model is performing more accurate with score of 84% in training set and 82% in test set , Thus Ridge Regression model will be the best choice for this assignment.

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 predictor variable in lasso area as follow :

GrLivArea	0.177071
BsmtQual_Ex	0.080819
Fireplaces	0.050871
GarageArea	0.048080
OverallQual_8	0.027147

House Price Prediction – Advanced Regression Assignment Part 2

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

A model needs to come built strong and generalizable so
that they are not impacted by outliers in the preparation dossier. The model endure further be generalisable because the test veracity is not inferior than the preparation score. The model should be correct for datasets apart
from the one that were secondhand all the while preparation. Too
much weightage concede possibility not given to the outliers for fear
that the veracity called for one model is extreme. To guarantee that this is not
the case, the exception analysis needs expected finished and only those that are hav ing to do
with the dataset need expected employed. Those outliers that it does not conform to keep must be detached from the dataset. This would help increase the veracity of t he indicators created bythe model. Confidence breaks maybe secondhand (usually 3-5 standard deviations). This would help similar the prognoses fashioned apiece m odel. If the model is not strong , it cannot be trustworthy for predicting analysis.