International Institute of Information Technology Bangalore

Housing price prediction :
Assignment Submission

Prepared and submitted by

Krishna Murthi B

Course : Executive PG Programme in Machine Learning & AI Feb 2022
Date : 06-Jul-2022

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 alpha I got for Ridge and Lasso was 0.8 and 20 respectively.

So doubling them, alphas as 1.6 and 40.

The results are as follows

```
The top 10 features
```

Out[246]:

	Ridge	Lasso	Ridge1 - With double alpha	Lasso1 - With double alpha
LotArea	41608.114491	42828.771618	38442.499442	38835.616374
OverallQual	108753.861739	112702.245000	105671.654308	116107.506039
OverallCond	36909.135548	37962.970141	34907.576896	37172.926945
BsmtFinSF1	50465.496663	49302.791343	51216.650359	49801.906712
TotalBsmtSF	110265.090130	125305.906880	98429.970399	118886.318101
1stFlrSF	56382.361354	0.000000	59228.539312	0.000000
2ndFlrSF	40052.709903	14585.368753	36712.281469	10666.657051
GrLivArea	78407.530790	128363.568530	79390.798284	132510.056063
BedroomAbvGr	-49923.892355	-54864.454182	-43108.164952	-49838.901633
TotRmsAbvGrd	51593.918351	51629.283653	50856.708256	50398.974443

Assignment-based Subjective Questions

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?

Ridge

- R2 - train 0.8801372522634662
- R2 - Test 0.891584635351981

Lasso

- R2 - train 0.880346692786556
- R2 - Test 0.8923457114312279

We will use Lasso regression to solve the problem. This is based on the R2 values.

Assignment-based Subjective Questions

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?

Observation

Observation

- The R2 values is reduced from 0.88 to 0.82

The top 10 features are

- 1stFlrSF
- 2ndFlrSF
- GrLivArea
- BedroomAbvGr
- TotRmsAbvGrd
- GarageCars
- AgeofProperty
- MSZoning_FV
- MSZoning_RH
- MSZoning_RL

Assignment-based Subjective Questions

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?

In my understanding, a model should have the following characteristics…

a) Simpler model : the model should be simple

b) Not overfitting : The model should not rely too much on the data that is used for train data.

b) Robust : the model should work well on the unseen data; in other words it should be consistently accurate on the test data as well. This is where I want to emphaisise on the outliers. We need to keep the outliers which are needed and relevant to the business data. However outlier data which is not needed should be removed.

c) It should show low bias and low variance

Implication:

a) If a model is not robust then we will not able to rely on the prediction.

b) Since we don't know how the model will behave on an unseen data and if the data contains outliers the model should might behave weird.

c) Afterall the models are built to be tested on the test data. Unless the model is robust, we should not rely on the prediction of the model.

d) So it is very important that the models should be robust, product accurate results

Thank you

Photo by Phil on Unsplash