

House Pricing Advance Regression Subjective Questions

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 :

The optimal value of alpha for ridge and lasso regression are as below

Ridge	4
Lasso	0.0001

We are doubling the alpha values as below

Ridge	8
Lasso	0.0002

After the changes these are the most important predictors variables which will implement in house pricing with coefficients

	New Ridge	New Lasso
MSZoning_RL	0.118034	0.152019
GrLivArea	0.106785	0.122905
MSZoning_RM	0.093093	0.110392
OverallQual	0.070565	0.07337
TotalBsmtSF	0.05681	0.07002
MSZoning_FV	0.055052	0.056709
OverallCond	0.045345	0.044696
Foundation_PConc	0.031511	0.037459
MSZoning_RH	0.029046	0.03133
GarageCars	0.028954	0.02996

Question-2:

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:

The optimal lambda value in case of Ridge and Lasso is as below:

Ridge	4
Lasso	0.0001

The lambda value of ridge is greater than lasso

The r2_score of ridge is slightly higher than lasso for the test dataset as below .

Metric	Ridge Regression	Lasso Regression
R2 Score (Train)	0.919178	0.919414
R2 Score (Test)	0.899918	0.899581
RSS (Train)	9.886205	9.85727
RSS (Test)	5.455734	5.474126
MSE (Train)	0.010734	0.010703
MSE (Test)	0.013777	0.013824

So we will choose ridge regression to solve this problem

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:

Currently the 5 most important predictors are available in lasso model as below

1. MSZoning_RL
2. MSZoning_RM
3. GrLivArea
4. MSZoning_FV
5. OverallQual

As per the question after removing the important predictors we have created another module in lasso and we got the below new 5 predictor variables with the highly correlated

TotalBsmtSF	0.092219
FullBath	0.083168
HalfBath	0.069498
OverallCond	0.054315
FireplaceQu	0.04717

These are the new predictor variables which will use for house pricing

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:

- A model needs to be made robust and generalizable so that they are not impacted by outliers in the training data.
- The model should also be generalisable so that the test accuracy is not lesser than the training score.
- The model should be accurate for datasets other than the ones which were used during training.
- Too much weightage should not be given to the outliers so that the accuracy predicted by the model is high.
- To ensure that this is not the case, the outlier analysis needs to be done and only those which are relevant to the dataset need to be retained.
- Those outliers which it does not make sense to keep must be removed from the dataset.
- This would help increase the accuracy of the predictions made by the model.
- Confidence intervals can be used (typically 3-5 standard deviations).
- This would help standardize the predictions made by the model.
- If the model is not robust, it cannot be trusted for predictive analysis.