

Subjective Questions

Assignment Part-I

Question 1

Which variables are significant in predicting the price of a house ?

Answer:

Top 30 significant variable in predicting price of house is as below.

GrLivArea
OverallQual
BsmtFinSF1
Neighborhood_StoneBr
YearBuilt
ExterQual_TA
OverallCond
TotalBsmtSF
Neighborhood_NridgHt
GarageCars
ExterQual_Gd
LotArea
Neighborhood_NoRidge
SaleType_New
Neighborhood_Crawfor
MSSubClass
Functional_Typ
BsmtExposure_Gd
KitchenQual_TA
ExterQual_Fa
KitchenQual_Gd
KitchenQual_Fa
LotFrontage
Street_Pave
MasVnrArea
BsmtQual_Gd
MSZoning_FV
Exterior1st_BrkFace
Condition1_Norm

Neighborhood_BrkSide

Question 2

How well those variables describe the price of a house. ?

Answer:

The R2 score of both Ridge and Lasso is as below.

Score	Ridge	Lasso
R2 Score (Train)	.93	.92
R2 Score (Test)	.92	.92

Assignment Part-II

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 as below

Ridge: 3

Lasso: 100

If we double the value of alpha for both Ridge and Lasso, we have following observations

- Ridge : The R2 score of training data as well testing data has decreased slightly
- Lasso : The R2score of training data as well testing data has decreased slightly.

The top 5 important predictor variables after this changes are as below

GrLivArea
OverallQual
BsmtFinSF1
Neighborhood_StoneBr
YearBuilt

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:

The consistency of the r2_score for lasso is evident in both the training and test datasets, making lasso the preferred choice in this scenario. Lasso aids in feature reduction within the model, ultimately leading to a more straightforward final model. This aspect is crucial for establishing a resilient and widely applicable model, as elaborated upon in the query.

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:

The top5 predictor after removing is as below.

TotalBsmtSF
2ndFlrSF
GarageCars
ExterQual_TA
FullBath

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:

- The model must possess generalization capabilities to ensure that the test accuracy is equal to or greater than the training score.
- It is crucial for the model to accurately predict outcomes for datasets that were not part of the training process.
- While outliers should not be disregarded entirely, their influence on the model's accuracy should be minimized.
- To achieve this, a thorough analysis of outliers must be conducted, and only those that are relevant to the dataset should be retained.
- Any outliers that do not contribute meaningfully to the dataset should be eliminated. If the model lacks robustness, it cannot be relied upon for predictive analysis.