

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?

Based on the Final model the optimal value of Alpha is

- Lasso Regression – 0.001
- Ridge Regression – 6

After Doubling the value of R2 is getting reduced for both ridge and lasso but not much significant change

	Alpha	R2 Score (Train)	R2 Score (Test)
Lasso Regression	0.001	0.931264	0.893139
	0.002	0.916045	0.875397
Ridge Regression	6	0.951020	0.902448
	12	0.9461387	0.902057

Most Important Predictors in order after the change

Lasso Regression	Ridge Regression
GrLivArea	GrLivArea
OverallQual_9	OverallQual_9
OverallQual_8	OverallQual_8
KitchenAbvGr_1	Neighborhood_Crawfor
Neighborhood_Crawfor	OverallCond_9
OverallQual_7	TotalBsmtSF
Functional_Typ	Functional_Typ
TotalBsmtSF	Neighborhood_Somerst
Condition1_Norm	OverallCond_8
OverallCond_7	Condition1_Norm

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?

The optimal value of Alpha is

- Lasso Regression – 0.001
- Ridge Regression – 6

Here the number of parameters selected by Lasso is 75 and that of Ridge is 276

As the problem statement is to find the most important feature for target variable, Lasso will be the better choice. Because the ridge tries to getting involved all the columns as much as possible which suppressed the effect of the most important variable.

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?

The Most Important feature for lasso are

Lasso Regression
GrLivArea
OverallQual_9
OverallQual_8
KitchenAbvGr_1
Neighborhood_Crawfor

The most important feature after removing the above variable is

Lasso Regression
OverallCond_9
OverallQual_10
SaleType_ConLD
OverallQual_3
Neighborhood_MeadowV

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?

To make the model Robust there should be very minimal impact of outlier in the training set. The Outlier treatment or removal of the value which is more deviated from the dataset should be performed to increase the accuracy of the model.

In order to make the model generalisable the test accuracy is not lesser than the training score which is the model should be accurate for datasets other than training set as well

Bias- Variance trade-off

Model with high bias pay very little attention to training data on the other hand model with high variance pays lot of attention to training data. So, in general we need to find the good balance without overfitting/underfitting the dataset to make the model more robust and generalisable.