

Advanced Regression-Subjective Questions and Answers

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

The Optimal Values of alpha for ridge and lasso are 0.00079901 & 0.00058701.

If we choose to double the value for both ridge and lasso then there will be a reduction in the test accuracy & Variance and an increase in the test error and bias.

The most important predictor variable is the coefficients for the columns given below.

Columns	Coefficients(-Ve)	Columns	Coefficients(+Ve)
BsmtQual_Gd	-0.2014	Functional_Typ	0.2584
GarageType_Basment	-0.1671	GrLivArea	0.4054
KitchenQual_Fa	-0.1841	Neighborhood_Crawfor	0.3317
KitchenQual_Gd	-0.243	Neighborhood_StoneBr	0.4514
KitchenQual_TA	-0.2195	SaleType_New	0.2537

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

	Ridge	Lasso
MSE Test	0.090746	0.072316
MSE Train	0.064382	0.059215
R2 Test	0.867015	0.894023
R2 Train	0.918182	0.924749
RSS Test	34.574249	27.55246
RSS Train	57.107245	52.523712

As can be observed, the Lasso performed better than the Ridge and had higher accuracy on the test (89>86) with less MSE.

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-predictor variable is given below

Functional_Sev	0.6461
GarageQual_TA	0.4497
HouseStyle_2.5Fin	0.9349
MSZoning_RH	0.4671
SaleCondition_Alloca	0.4953

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

When the model has low variance and low bias it is said to be robust. The model should be simple
When we make the simpler model the training accuracy is reduced and test accuracy is dependent on variance and bias.