

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 –

Optimal value of alpha for ridge regression is 7 and for lasso regression it is 100.

After doubling the alpha for ridge and lasso regression, R2 score of the models decreased for training data while RSS and MSE increased for training data.

Top 5 predictor variables in Ridge regression

Before Change	After Change
2ndFlrSF	Neighborhood_NoRidge
Neighborhood_NoRidge	2ndFlrSF
RoofMatl_WdShngl	OverallQual_9
OverallQual_10	OverallQual_10
OverallQual_9	GrLivArea

Top 5 predictor variables in Lasso regression

Before Change	After Change
Condition2_PosN	GrLivArea
PoolQC_Gd	OverallQual_10
Exterior2nd_Stucco	Condition2_PosN
GrLivArea	OverallQual_9
OverallQual_10	RoofMatl_WdShngl

Description of each of the above-mentioned predictor variables is as follows

2ndFlrSF - Second floor square feet

Neighborhood_NoRidge – Northridge in the neighborhood

RoofMatl_WdShngl – Roof material is Wood Shingles

OverallQual_10 – Overall material and finish of the house is “Very Excellent”

OverallQual_9 – Overall material and finish of the house is “Excellent”

GrLivArea – Above grade (ground) living area square feet

Condition2_PosN – Proximity to various conditions - Near positive off-site feature

PoolQC_Gd – Pool quality – Good

Exterior2nd_Stucco – Second exterior covering on house - Stucco

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 –

I would choose to apply Lasso regression for building the model. There are 282 independent variables/features available after data preparation stage.

Ridge regression will make use of all 282 features to build the model which will make the model quite complex. Hence, to keep the model simple, Lasso regression is the right way forward.

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 next 5 most important predictor variables are

- 1) Roof material - Wood Shingles
- 2) Overall material and finish of the house - "Excellent"
- 3) Neighborhood – Northridge
- 4) LotArea - Lot size in square feet
- 5) GarageCars - Size of garage in car capacity

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 –

By selecting appropriate value of lambda in Ridge/Lasso regression we can ensure that the Variance as well as Bias of the model is low, which ensures that the model is robust and generalizable.

Accuracy of the model improves when appropriate lambda is selected as it would mean low bias and low variance, which in-turn translates to better prediction and thus better accuracy of the model.