
Q1) 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?

Ans: Value of parameter alpha is as below :

Ridge: For ridge it is .9

Lasso: For lasso it is 50

In general, if we increase the value of alpha it means we want to have high value of regularization and hence want to reduce overfitting and remove noise

Below are top 5 factors after doubling :

1) Total square feet of basement area [TotalBsmtSF]

2) overall material and finish of the house[OverallQual]

3) Roof material[RoofMatl]

4) Size of garage in car capacity[GarageCars]

5) Physical locations within Ames city limits[Neighborhood_NoRidge]

Q2) **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?

Ans: Both Lasso and Ridge ha done good on train and test data , after lasso R-square value on test data set was slightly hence I would go for Lasso , also some coefficients were made 0 for model simplicity it would be good.

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

ANS: Other variables would be :

1) Overall condition of the house[OverallCond]
2) Rooms
3) Quality of the material on the exterior[ExterQual_TA]
4) Physical locations within Ames city limits[Neighborhood_Crawfor]
5) Quality of the material on the [ExteriorExterQual_Gd]

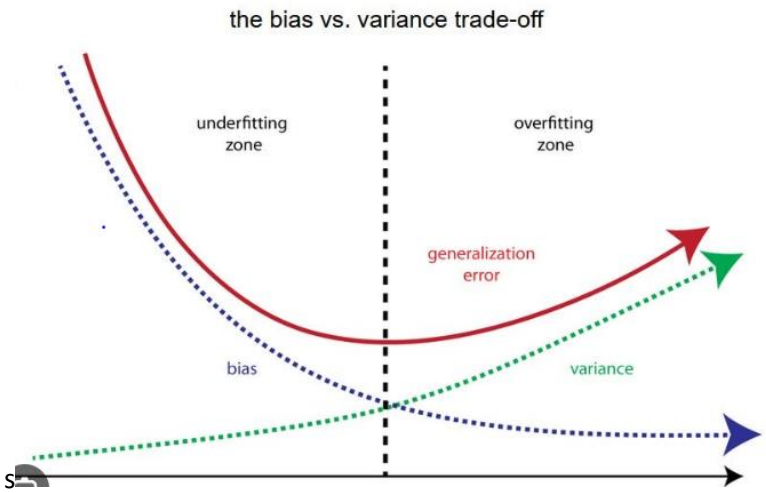
Question 4

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?

Model would be robust and generalisable if in case its performance on tets data is almost similar or close to how it performed on training data and so that means there is no overfitting done in the model and so variance is low

We also would need to make sure that no underfitting is done this can be identified if model Is not performing that good on both train and test data

We need to basically consider simple model(with less no of feature and complexity) which performs /qualifies good in bias variance trade-off


the bias vs. variance trade-off
underfitting zone
overfitting zone
generalization error
bias
variance