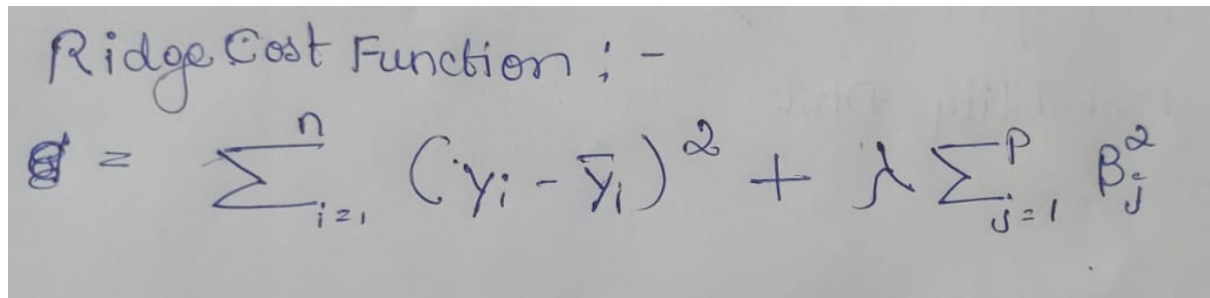


Question 1

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:

In Ridge regression uses the penalty to regularize i.e

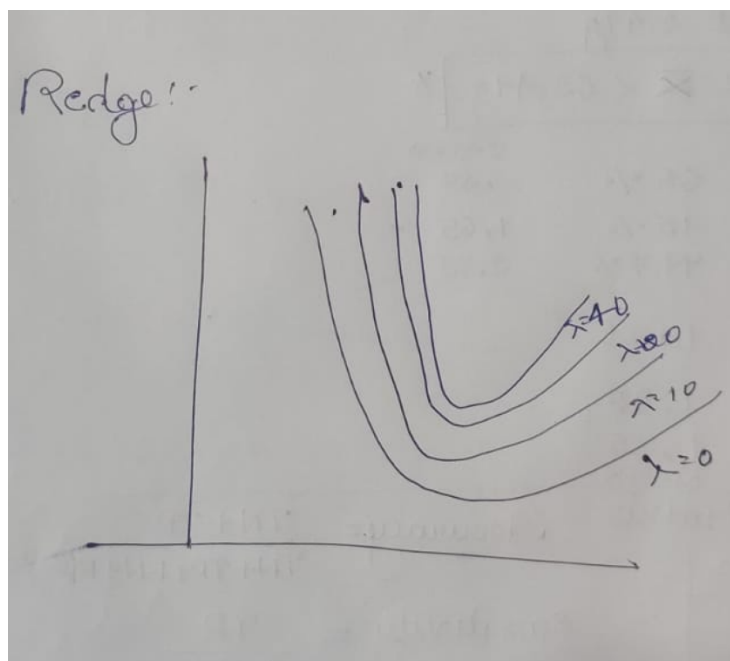


Ridge Cost Function :-

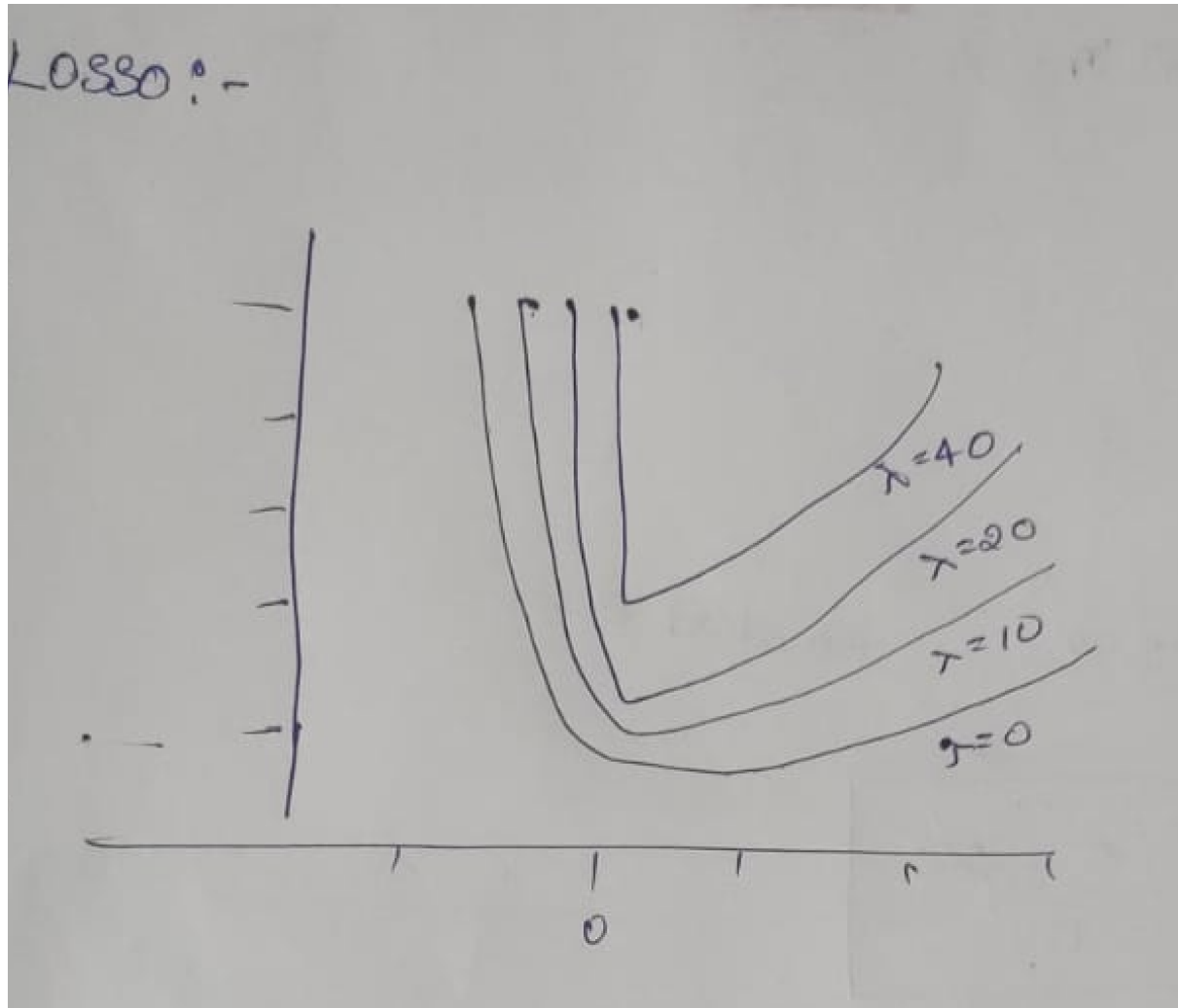
$$J = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^p \beta_j^2$$

The optimal value for the alpha is something which solves the overfitting and able to understand the pattern in the data and should not underfit as well. Reduce the variance by compromising on bias. If alpha value is too high then will lead to more regularization.

In Ridge regression When we increase the penalty(alpha) value, the optimal slope will shift to zero, As and when the alpha value will raised the optimal slope will be closer to zero but will never be zero.



In Lasso regression, When we increase the penalty(alpha) value, the optimal slope will move towards zero, As and when the alpha value will raised the optimal slope will be closer to zero and for the highest value of alpha the optimal slope will reach zero.



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:

Ridge:

optimal value for lambda is 20

Lasso:

Optimal value for lambda is 0.001

Both are the optimum minimal value for lambda tried to increase the value, as and when the value of lambda is increase it was leading more regularization as I could see the variance is increased.

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:

TotRmsAbvGrd, BedroomAbvGr, KitchenAbvGr, BsmtCond, BsmtFinType1, Heating_GasA

Question 4

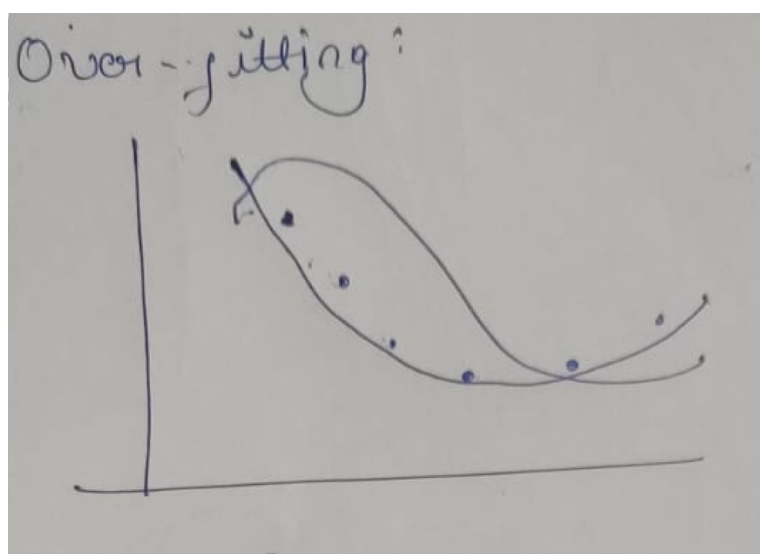
How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?

Answer:

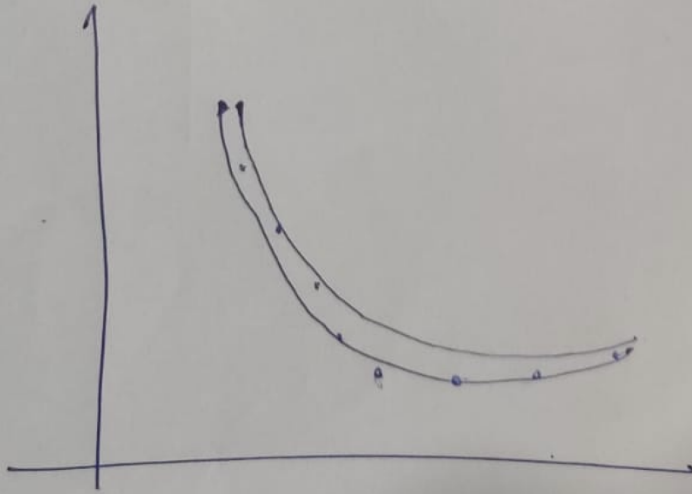
Model should perform well with both training data and test data, we can make the model more robust and generalised by reducing the variance I,e reducing the square error term,

if model accuracy is not good with both training and test data it will result in the underfitting.

if the model accuracy is good with training data and not good with testing data then it will lead to overfitting. Basically the Bias will be low but the variance will be very high



Best-fits -



Under-fitting

