

Both this document and python file also uploaded to -
<https://github.com/KunalBhosale/AdvancedRegression>

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?

Ans:

- Optimal value of **alpha Ridge - 0.1, Lasso – 50**
- Once double value of alpha was chosen, **r2 score increased to 73.82 for Ridge, and 73.93** for Lasso, also coefficients of all the predictor variables changed, some increased, and for some it decreased
- 9 Most important predictor variables after change:

Ridge: GrLivArea, RoofMatl_WdShngl, 1stFlrSF, 2ndFlrSF, TotalBsmtSF, LotArea, BsmtFinSF1, RoofMatl_Membran, OverallQual

Lasso: GrLivArea, TotalBsmtSF, OverallQual, RoofMatl_WdShngl, Neighborhood_NoRidge, OverallCond, GarageCars, Lot area, BsmtFinSF1

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: I will choose to apply **Ridge – 0.2 and Lasso – 100**, because with these lambda values R2 score of the model increased. Both Ridge and Lasso are having almost same R2 score on test data, I will go with Lasso model.

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:

Could not try this option, but I think most important predictor variables would be the next 5 predictor variables

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?

Ans:

To make model robust and generalisable, model should not be very complex, it should not overfit nor underfit. It should be able to identify most common patterns in the train data set, and then should be able to identify general patterns in the test dataset with low variance but there could be small bias.

So in short model should have low variance and low bias.