

# Advanced Regression Assignment

## Part – 2

- 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) The optimal value of alpha for ridge and lasso regression are 7 and 100. If we double the value of alpha for ridge, then train  $r^2$  increases and test  $r^2$  approximately same. However

| Important variable with alpha 7 | Important Variable with alpha 14 |
|---------------------------------|----------------------------------|
| GrLivArea                       | LotFrontage                      |
| OverallQual                     | LotArea                          |
| 1stFlrSF                        | OverallQual                      |
| TotBsmtSF                       | OverallCond                      |
| LotArea                         | MasVnrArea                       |

Now if we double the value of alpha for lasso from 100 to 200 test  $r^2$  decreases and train  $r^2$  increases.

| Important Variable with alpha 100 | Important Variable with alpha 200 |
|-----------------------------------|-----------------------------------|
| GrLivArea                         | LotFrontage                       |
| OverallQual                       | LotArea                           |
| Condition2_PosN                   | OverallQual                       |
| OverallCond                       | OverallCond                       |
| LotArea                           | MasVnrArea                        |

All of this calculations are attached in jupyter notebook in section named part2.

**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) We will choose Lasso because of these two reasons

- a) Every metric is slightly better in lasso than ridge.
- b) Lasso make many variables coefficient zero, so we get to know model interpretability very well.

**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) We created another model excluding five important Variable. Then the 5 important variables are

- a) 2ndFlrSF
- b) 1stFlrSF
- c) TotBsmtSF
- d) Neighborhood\_NridgHt
- e) Neighborhood\_NoRidge

The last two variable are one hot encoded variable of neighbourhood column.

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) If a model is robust and generalisable it means it correctly understand the data pattern, nor it is overfit(learning noise) neither it is Underfit(learning nothing). For such a model bias and variance will be low and it will perform good on unseen data as well. If a model is robust then train accuracy is good and test accuracy also drop very slightly because it is perfect fit and neither its bias is high nor variance as we have done the perfect tradeoff between bias and variance.