

## 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:

From the analysis we can infer that optimum value of alpha for Ridge regression is 4 with train R2 score as 0.9328774732283216 and test data 0.8896604150718403 and for alpha is 0.0001 with train R2 score as 0.9439879231554645 and test data 0.8965367999723738.

After checking coefficients when doubled alpha to 0.0002 for lasso and 8 for Ridge below are top features for the both regressions.

Lasso regression top 5 important predictor variables after double the value of alpha is

```
GrLivArea
1stFlrSF
TotalBsmtSF
2ndFlrSF
LotArea
```

Ridge regression top 5 important predictor variables after double the value of alpha is

```
GrLivArea
MSZoning_FV
TotalBsmtSF
gap_between_build_remodel
MSZoning_RL
```

## 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:

By seeing the results, we can see Ridge model got to explain 93.2% on train data and 88.96% with test data variations which is pretty good model.

Lasso model got to explain 94.39% on train data and 89.65% with test data variations which is a pretty good model and we got feature elimination by seeing coefficients moving to zero used for feature elimination.

Both models are pretty in terms of R2score on train data and test data.

As we are getting feature elimination in lasso, we can say lasso is best suited in this case as both have similar R2 score.

### 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:

From lasso model we can see top 10 features if top 5 features are eliminated then we will be left with the

- LotArea
- BsmtFinSF1
- GarageArea
- Neighborhood\_Crawfor
- gap\_between\_build\_remodel

From Ridge model after eliminating top 5 features, we will be left with are

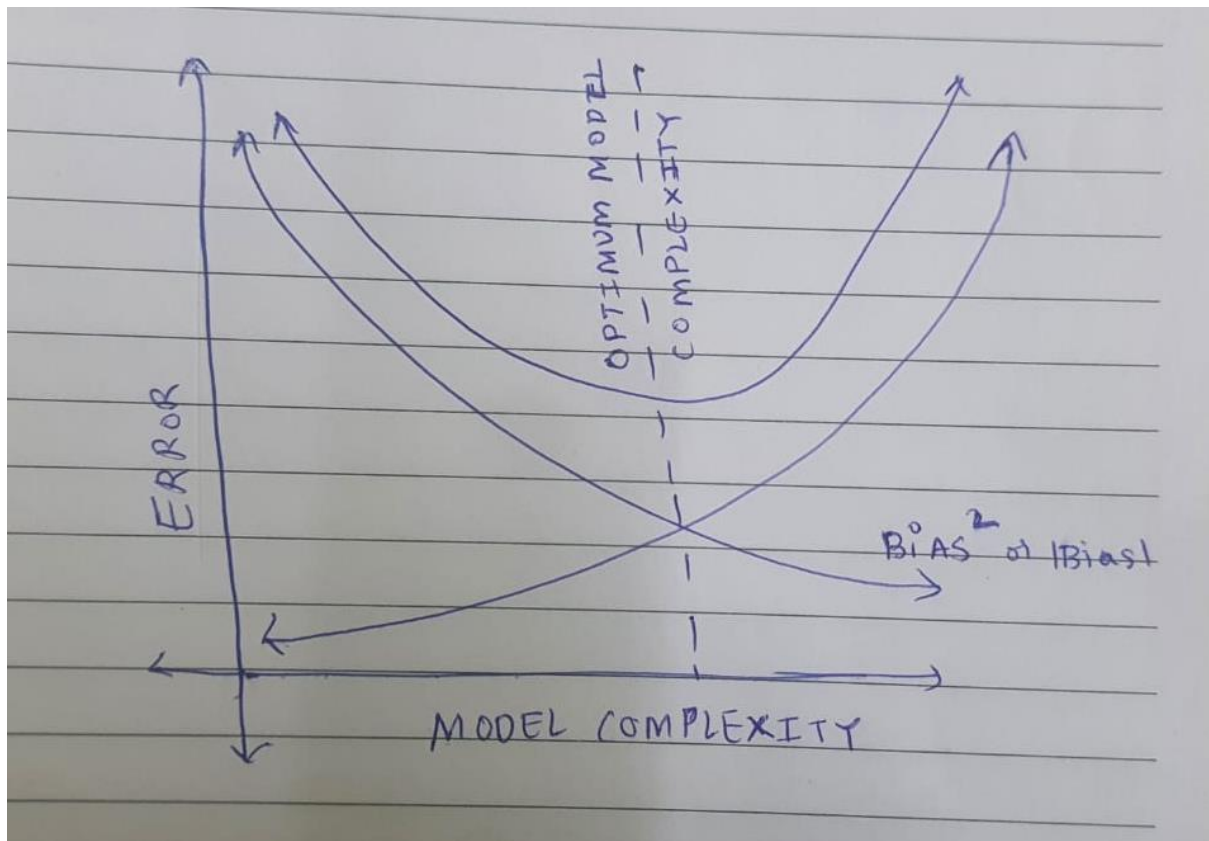
- MSZoning\_RL
- Condition2\_PosA
- Neighborhood\_Crawfor
- MSZoning\_RM
- GarageArea

### 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:** The model should be as straightforward as feasible because this will increase its robustness and generalizability while reducing accuracy. The Bias-Variance trade-off can also be used to understand it. The bias increases with model complexity while decreasing variance and increasing generalizability. Its accuracy implication is that a robust and generalizable model will perform similarly on both training and test data, i.e., the accuracy does not change significantly for training and test data.



**Bias:** Bias is a model error that occurs when the model cannot learn from the data. High bias prevents the model from understanding the nuances of the data. On training and testing data, the model performs poorly.

**Variance:** When a model tries to overlearn from the data, variance occurs. High variance indicates that the model performs remarkably well on training data since it was well trained on that data, but it performs dreadfully on testing data because it was uncharted territory for the model.

To prevent the overfitting and underfitting of data, bias and variance must be in balance.