Questions & Answers

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: Below are the R2 score of ridge and lasso regression before doubling the alpha values:

- Alpha value for ridge regression is 0.01
  Train R2 score: 0.8515575896823354
  Test R2 score: 0.8182274393837714
- Alpha value for lasso regression is 0.0001
  Train R2 score: 0.822500680561198
  Test R2 score: 0.8325534134606531

Below are the R2 score of ridge and lasso regression after doubling the alpha values:

- Alpha value for ridge regression is 0.02
  Train R2 score: 0.849486098450797
  Test R2 score: 0.824069688634385
- Alpha value for lasso regression is 0.0002
  Train R2 score: 0.8160483762349595
  Test R2 score: 0.8223987644522853

The alpha values are doubled we can see that R2 values are decreased.
The important predictor variables are:

- GrLivArea
- Exterior1st_BrkComm
- TotalBsmtSF
- ExterQual_Fa
- KitchenAbvGr
- Functional_Maj2
- GarageCars
- LotArea
- Functional_Sev
- ExterQual_TA

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: ridge regression:

Train R2 score: 0.8515575896823354
Test R2 score: 0.8182274393837714

lasso regression:

Train R2 score: 0.822500680561198
Test R2 score: 0.8325534134606531

- The Ridge regression r2 score is greater than the lasso regression r2 score.
- The model becomes complex when there are many features but in lasso regression it pushes some of the coefficients which are no necessary exactly to 0.

Hence, Variables predicted by lasso regression can be applied for predicting the price of a house.

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: After removing the five most important predictor variables in lasso model.
The most important predictor variables are as follows:

- LotFrontage
- HouseStyle_2.5Fin
- HouseStyle_2.5Unf
- Total-porch_sf
- Condition2_PosA

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:

- Simpler models are usually more generic and are more widely applicable
- Simpler models are more robust. Complex models tend to change wildly with the changes in the training data set simple models have low variance, high bias and complex models have low bias, high variance.
- Simpler models require fewer training samples for effective training than the more complex ones and hence are easier to train.
- Complex model will need to change for every little change in the dataset and hence is very unstable and extremely sensitive to any changes in the training data. A simpler model that abstracts out some pattern followed by the data points given is unlikely to change wildly even if more points are added or removed. Bias quantifies how accurate the model is likely to be on test data. A complex model can do an accurate job prediction provided there is enough training data. Variance refers to the degree of changes in the model itself with respect to changes in the training data. Thus accuracy of the model can be maintained by keeping the balance between bias and variance.