Problem Statement – Part II

Assignment Part-II

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:

The optimal value of alpha for lasso and ridge are 0.00001 and 1 respectively. If we increase the alpha (hyper parameter value) the accuracy of the model starts dropping gradually. It might increase a bit to the optimal hyper parameter value, but the accuracy will decrease with the increase in alpha and model will become more parse.

The most important predictor variables before and after the change is implemented is as follows:

Ridge Regression

Features that ridge provides,

at optimal alpha (1):

	ridge coefficient
Total_sqr_footage	0.085296
GarageArea	0.058171
TotRmsAbvGrd	0.044423
LotArea	0.035191
Neighborhood_StoneBr	0.022221
Neighborhood_Crawfor	0.022186
LotFrontage	0.021812
SaleType_CWD	0.020355
HouseStyle_2.5Unf	0.020233
OverallCond	0.018852
Total_porch_sf	0.018835
RoofMatl_WdShngl	0.017610
RoofStyle_Shed	0.017175
Neighborhood_NoRidge	0.017025
Exterior1st_BrkFace	0.016305
ScreenPorch	0.015496
CentralAir_Y	0.014873
Functional_Typ	0.014179
Condition2_PosA	0.013500
Neighborhood_BrkSide	0.013286

Features that ridge provides.

at double the optimal alpha (2):

	ridge coefficient
Total_sqr_footage	0.080017
GarageArea	0.055136
TotRmsAbvGrd	0.043693
LotArea	0.031008
Neighborhood_StoneBr	0.021380
Neighborhood_Crawfor	0.021366
Total_porch_sf	0.019263
LotFrontage	0.018581
HouseStyle_2.5Unf	0.017917
OverallCond	0.017571
Neighborhood_NoRidge	0.017426
SaleType_CWD	0.016870
Exterior1st_BrkFace	0.016411
RoofMatl_WdShngl	0.015797
CentralAir_Y	0.015247
ScreenPorch	0.015223
Functional_Typ	0.013459
Neighborhood_Somerst	0.012888
Neighborhood_Veenker	0.012545
Neighborhood_BrkSide	0.012262

Lasso Regression

Features that lasso provides,

at optimal alpha (0.00001):

	lasso coefficient
Total_sqr_footage	0.091416
GarageArea	0.061726
TotRmsAbvGrd	0.044693
LotArea	0.040584
RoofStyle_Shed	0.035967
HouseStyle_2.5Unf	0.022576
SaleType_CWD	0.022452
LotFrontage	0.022425
Neighborhood_Crawfor	0.022273
Neighborhood_StoneBr	0.021658
OverallCond	0.018698
RoofMatl_WdShngl	0.018550
Total_porch_sf	0.017773
RoofMatl_Metal	0.016458
ScreenPorch	0.015726
Neighborhood_NoRidge	0.015693
Exterior1st_BrkFace	0.014939
Functional_Typ	0.014446
CentralAir_Y	0.014445
Neighborhood_BrkSide	0.014332

Features that lasso provides.

at double the optimal alpha (0.00002):

	lasso coefficient
Total_sqr_footage	0.091831
GarageArea	0.061586
TotRmsAbvGrd	0.044721
LotArea	0.038024
Neighborhood_Crawfor	0.021773
HouseStyle_2.5Unf	0.021658
Neighborhood_StoneBr	0.021228
SaleType_CWD	0.019348
LotFrontage	0.018232
OverallCond	0.018226
Total_porch_sf	0.017710
RoofStyle_Shed	0.016871
RoofMatl_WdShngl	0.016577
Neighborhood_NoRidge	0.015243
ScreenPorch	0.015103
Exterior1st_BrkFace	0.014822
Functional_Typ	0.014591
CentralAir_Y	0.014550
ExterCond_Ex	0.013528
Neighborhood_BrkSide	0.013512

When we double the value of alpha for our ridge regression the model will apply more penalty on the curve and try to make the model more generalized that is making model simpler and no thinking to fit every data of the data set. Similarly, when we increase the value of alpha for lasso, we try to penalize more our model and more coefficient of the variable will be reduced to zero, when we increase the value of our r2 square also decreases.

Overall, since the alpha values are small, we do not see a huge change in the model after doubling the alpha.

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:

- The optimum lambda value in case of Ridge and Lasso is as follows: -

- Ridge - 1
- Lasso - 0.00001

- The Mean Squared Error in case of Ridge and Lasso are:
  - Ridge - 0.0005749363811384514
  - Lasso – 0.0005836113760506755
- The R2 Score Value in case of Ridge and Lasso are:
  - Ridge - 0.852148945901396
  - Lasso - 0.8499180779581766

The Mean Squared Error & R2 score of both the models are almost same. Since Lasso helps in feature reduction (as the coefficient value of some of the features becomes zero) & will penalize more on the dataset, Lasso has a better edge over Ridge and should be used as the final 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?

Answer:

The five most important predictor variables in the current lasso model is:

	lasso coefficient
Total_sqr_footage	0.091416
GarageArea	0.061726
TotRmsAbvGrd	0.044693
LotArea	0.040584
RoofStyle_Shed	0.035967

We build a Lasso model in the Jupyter notebook after removing these attributes from the dataset.

The R2 of the new model without the top 5 predictors drops to 0.777346362375955.

The Mean Squared Error increases to 0.0008658151099653831.

The new Top 5 predictors are:

	lasso coefficient
LotFrontage	0.080901
HouseStyle_2.5Unf	0.046354
RoofMatl_WdShngl	0.039407
Total_porch_sf	0.039136
Neighborhood_NoRidge	0.039121

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:

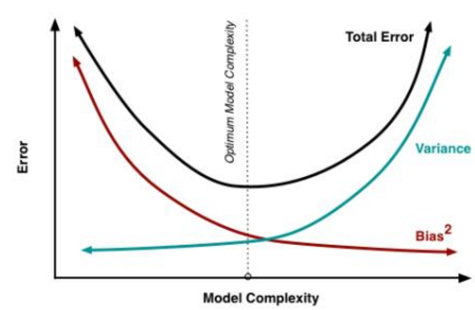
As Per, Occam's Razor— given two models that show similar 'performance' in the finite training or test data, we should pick the one that makes fewer on the test data due to following reasons:

- Simpler models are usually more 'generic' and are more widely applicable.
- Simpler models require fewer training samples for effective training than the more complex ones and hence are easier to train.
- Simpler models are more robust.
    - Complex models tend to change wildly with changes in the training data set.
    - Simple models have low variance, high bias and complex models have low bias, high variance.
    - Simpler models make more errors in the training set. Complex models lead to overfitting — they work very well for the training samples, fail miserably when applied to other test samples.

Therefore, to make the model more robust and generalizable, make the model simple but not simpler which will not be of any use. Regularization can be used to make the model simpler. Regularization helps to strike the delicate balance between keeping the model simple and not making it too naive to be of any use. For regression, regularization involves adding a regularization term to the cost that adds up the absolute values or the squares of the parameters of the model.

Also, Making a model simple lead to Bias-Variance Trade-off:

- A 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 is the model likely to be on test data. A complex model can do an accurate job prediction provided there is enough training data. Models that are too naïve, for e.g., one that gives the same answer to all test inputs and makes no discrimination whatsoever has a very large bias as its expected error across all test inputs are very high.

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 as it minimizes the total error as shown in the above graph.