Advance Linear Regression Assignment Subjective Question

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 alpha value obtained for ridge regression is **100** and for lasso it is **0.001**

Ridge Regression:

- **Model Build using alpha as 100 we obtained the following results**
  - ✓ For Ridge regression with alpha value at 100 R2_Score obtained for training and test data are **0.94 and 0.91** respectively
  - ✓ RMSE train and test data are **0.0017 and 0.0027** respectively
  - ✓ Most important predictor variables obtained were :

OverallQual	0.023
GrLivArea	0.021
1stFlrSF	0.016
OverallCond	0.015
TotalBsmtSF	0.013
BsmtFinSF1	0.011
LotArea	0.011
FullBath	0.011
2ndFlrSF	0.010
Neighborhood_StoneBr	0.009

- **Model Build using alpha as 200 we obtained the following results**
  - ✓ For Ridge regression with alpha value at 200 R2_Score obtained for training and test data are **0.93 and 0.91** respectively
  - ✓ RMSE train and test data are **0.0018 and 0.0027** respectively
  - ✓ Top 10 features obtained after building model :

OverallQual	0.020
GrLivArea	0.019
1stFlrSF	0.015
TotalBsmtSF	0.013
OverallCond	0.012
BsmtFinSF1	0.011
LotArea	0.010
FullBath	0.010
2ndFlrSF	0.009
KitchenQual	0.008

Lasso Regression:

- **Model Build using alpha as 0.001 we obtained the following results**
  - ✓ For Lasso regression with alpha value at 0.001 R2_Score obtained for training and test data are **0.93 and 0.92** respectively
  - ✓ RMSE train and test data are **0.0018 and 0.0024** respectively
  - ✓ Total number features selected by the model was **85**
  - ✓ Most important predictor variables obtained were :

GrLivArea	0.049
OverallQual	0.032
OverallCond	0.016

TotalBsmtSF	0.014
BsmtFinSF1	0.012
LotArea	0.012
GarageCars	0.011
MSZoning_RL	0.008
CentralAir	0.008
Neighborhood_Crawfor	0.008

- **Model Build using alpha as 0.002 we obtained the following results**

  - ✓ For Lasso regression with alpha value at **0.002** R2_Score obtained for training and test data are **0.93 and 0.91** respectively
  - ✓ RMSE train and test data are **0.0020 and 0.0025** respectively
  - ✓ Total number features selected by the model was **62**
  - ✓ Top 10 features obtained after building model :

GrLivArea	0.049
OverallQual	0.032
OverallCond	0.016
TotalBsmtSF	0.014
BsmtFinSF1	0.012
LotArea	0.012
GarageCars	0.011
MSZoning_RL	0.008
CentralAir	0.008

Neighborhood_Crawfor	0.008

Summary:

Overall the model accuracy seems to be the same. However for Ridge we can see considerable change in the coefficients after doubling the alpha parameter while for Lasso we can see considerable number of features were eliminated or were considered insignificant and we ended up with 62 features 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:

- ✓ After applying GridSearchCV, we arrived at the following alpha (hyper parameters):
    - **Ridge –** *100*
    - **Lasso –** *0.001*
- ✓ R2 score obtained were:
    - **Ridge** – *0.94 and 0.91 for the train and test data*
    - **Lasso** -- *0.94 and 0.92 for the train and test data*
- ✓ Root Mean Square score obtained were:
    - **Ridge** – *0.002657*
    - **Lasso** -- *0.002447*
- ✓ It can be observed that Lasso has slightly lesser RMSE compared to Ridge.
- ✓ Also Lasso provides additional feature of eliminating the features. It reduced total number of significant features to 85 by making their coefficients 0 instead of Ridge which doesn't make the coefficients of the features as 0

Summary: *Overall both Ridge and Lasso performed well by giving approximately same R2 and RMSE scores and in addition since Lasso eliminated features as well, which makes Lasso* *preferred one. Hence, variables chosen by Lasso can be applied to choose significant variable to predict the price of the house*

Question 3

After building the model, you realized 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:

- We dropped the following top 5 features to build the model again with 187 features:
  - ✓ GrLivArea
  - ✓ OverallQual
  - ✓ BsmtFinSF1
  - ✓ OverallCond
  - ✓ TotalBsmtSF
- After performing GridSearchCV we obtained the best hypermeter for this model is **0.001**
- R2 score obtained for train and test data **were 0.93 and 0.88** respectively which isn't bad. However the model before dropping these 5 features were giving us better R2 score.
- RMSE obtained was **0.0034**
- Total number features were now reduced to **93 from 187 features in the original dataset**
- Important features obtained after eliminating the top 5 features from the data set were:

1stFlrSF	0.057
2ndFlrSF	0.044
CentralAir	0.013
KitchenQual	0.012

BsmtFinType1	0.011

Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer:

According to **Occam's razor** given two models that perform similarly, with both finite test and training data, we should choose the one which has fewer features and is simpler and generalizable. It can further be explained by the following principals:

- Simpler models are more **generic** and widely applicable
- Simpler model require **fewer training samples** compared to the complex ones. This makes training simpler model easier with limited data
- Simpler model are **more robust.** Simpler model will not be sensitive to the specifics of the training data.
  - Complex models will change wildly with the change in the training data.
  - Simpler models will have high bias and low variance and complex model because of overfitting, might have low bias and high variance.
- Simpler model make **more errors in the training data set**. While complex models due to overfitting, they might work very well with the training data but fail miserably with the test data.

Therefore to make the model robust and generalizable, the model should be simple but not too naïve that it is useless.

Regularization:

Regularization can be used to make the models simpler by striking a balance between keeping the model simple but not making them too naïve to be of any use. By regularization, we add a regularization term to the cost that add up the absolute values or square of the absolute value to the error term.

Two main regularized regression models are:

- **Ridge** - In Ridge an additional term "sum of squares of coefficients" Is added to the cost function along with error term.

$$\min\left(\left||\mathrm{Y} - \mathrm{X}(\theta)\right||_2^2 + \lambda||\theta||_2^2\right)$$

- **Lasso –** In Lasso , "sum of absolute value of coefficients" is added to the cost function along with the error term

$$\min\left(\left||\mathrm{Y} - \mathrm{X}\theta\right||_2^2 + \lambda||\theta||_1\right)$$

By increasing the hyper parameter, ideally we penalize the coefficient there by making the model simpler and not too complex.

Bias- Variance Trade off

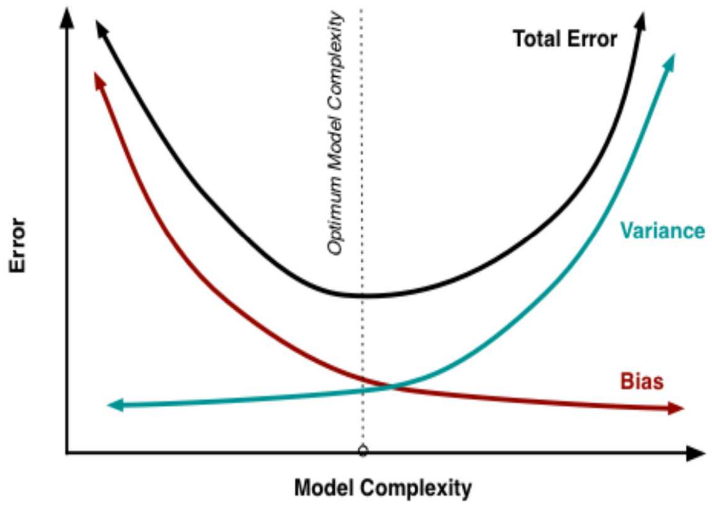
However, making the model simple, leads to bias variance trade-off.

- A complex model, will have to change for every change in training dataset. Hence very unstable and would be extremely sensitive to the change in training dataset
- A simpler model is very generic and any change in training data set would not have major impact if the training dataset changes.

Bias explains, how accurate the model performs on the test data. Complex model, assuming we have quantifiable training data available would perform very well. Models that are too naïve would not perform well. Models that are too naïve may give the same output irrespective what the input is and would make no discrimination.

Variance refers to the change in the model, with the change in the training data. If a model overfits, with change in data the model would not predict very well. This behavior can be mainly observed in the complex data.

Hence, a balance should be maintained between the Bias and Variance and reduce the error in the model by keeping the model simple but not too naïve.



Optimum Model Complexity
Total Error
Variance
Error
Bias
Model Complexity