Assignment Part -II - Subjective Questions and Answers

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 LAMBDA we got in case of Ridge and Lasso is :

  1. Ridge - **1.0**
  2. Lasso - **0.0001**

- The changes in the model if we choose to double the value of alpha for both ridge and lasso are:

  ➢ When we double the alpha value in case of ridge there is a slight increase in the mean squared error whereas the r2 value of train and test remains almost same.

  ➢ When we double the alpha value in case of lasso there is a slight increase in the mean squared error, the r2 value of train slightly decreases whereas there is a huge fall in the r2 value of test thus making the model and prediction worse.

  ➢ It also penalizes the model even more and a greater number of coefficients of a variable shrink towards zero.

- The most important predictor variables after the changes are implemented are:

  For Ridge:

  1. LotFrontage
  2. OverallCond
  3. BsmtUnfSF
  4. BsmtFinType2
  5. OverallQual
  6. BsmtFinSF1
  7. BsmtExposure
  8. GarageFinish
  9. LotConfig_CulDSac
  10. ExterCond

For Lasso:

1. LotFrontage
2. BsmtFullBath
3. OverallCond
4. CentralAir
5. OverallQual
6. Exterior1st_CBlock
7. MSZoning_RH
8. MSZoning_RM
9. Street_Pave
10. GarageQual

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 optimal value of LAMBDA we got in case of Ridge and Lasso is:

- Ridge - **1.0**
- Lasso - **0.0001**

The r2 value we got in case of Ridge and Lasso is:

- Ridge - Train = **0.927**, Test = **0.745**
- Lasso - Train = **0.928**, Test = **0.737**

The Mean Squared error in case of Ridge and Lasso is:

- Ridge - **0.00297**
- Lasso - **0.00280**

➢ We can clearly observe that the Mean Squared Error of Lasso is slightly lower than that of Ridge.

➢ Also, since Lasso helps in feature reduction (as the coefficient value of one of the lasso's features to be shrunk toward 0) and helps to increase model interpretation by taking the magnitude of the coefficients thus **Lasso has a better edge over Ridge. Hence we will** choose to apply Lasso because of the above stated reasons

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

We dropped the top 5 most important predictor variables in the lasso model and again created the model and got the below five most important predictor variables:(refer the python file)

1. LotArea
2. FullBath
3. 1stFlrSF
4. ExterCond
5. MSZoning_RH

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 is expected to be as simple as possible and simpler models are considered as more 'generic', though its accuracy will be decreased but it will be more robust.
- This can be understood from the Bias-Variance trade-off. The simpler the model the more the bias but less variance becoming generalizable. Whereas the complex model will have high variance and low bias.
- Sometimes underfitting and overfitting are the problems associated with the model. Hence, it is important to have balance in Bias and Variance to avoid such problems. This is possible with "Regularization".
- Regularization helps in managing the model complexity by essentially shrinking the coefficients towards zero. This avoids the model becoming too complex, thus reducing the risk of overfitting.
- Regularization method should be used to keep the model optimum simpler. It penalizes the model if it becomes more complex.
- Regularization method helps to achieve the Bias-Variance trade off. It compromises by increasing bias to a optimum position where Total Error is minimum.

- This point also known as Optimum Model Complexity where Model is sufficient simpler to be generalisable and also complex enough to be robust.

- Making a model simple lead to Bias-Variance trade off.