Subjective Questions

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:-

a. Optimal Values
- Ridge regression = 0.01
- Lasso regression = 0.001

b. If we doubled the alpha value
1. For Ridge it makes the model more generalized and more simpler which leads to over fitting of the model
2. For lasso it leads to penalize more our model and more coefficient of the variable will reduced to zero, when we increase the value of our r2 square also decreases

c. The most important predictor variables are

1. MSSubClass
2. RoofMatl_Membran
3. MSZoning_RL
4. MSZoning_FV
5. MSZoning_RH
6. MSZoning_RM
7. Condition2_PosA
8. RoofMatl_WdShngl

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: - I'll choose**,** Ridge regression, It uses a tuning parameter called lambda as the penalty is square of magnitude of coefficients which is identified by cross validation. Residual sum or squares should be small by using the penalty. The penalty is lambda times sum of squares of the coefficients, hence the coefficients that have greater values gets penalized. As we increase the value of lambda the variance in model is dropped and bias remains constant. Ridge regression includes all variables in final model unlike Lasso Regression.

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:-

1. GrLivArea
2. OverallQual
3. OverallCond
4. TotalBsmtSF
5. GarageArea

Subjective Questions

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:-

1. The model should be as simple as possible, though its accuracy will decrease but it will be more robust and generalizable.
2. It can be also understood using the Bias-Variance trade-off.
3. The simpler the model the more the bias but less variance and more generalizable.
4. Its implication in terms of accuracy is that a robust and generalizable model will perform equally well on both training and test data i.e. the accuracy does not change much for training and test data.