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 Ridge regression is **3.5** and for Lasso regression **0.001**

Below are the Metrics before and after alpha is double

Metric	**Ridge**	**Lasso**	**Ridge**	**Lasso**
			(Double Alpha)	
R^2 Train	0.9429	0.9382	0.9390	0.9314
MSE Train	0.0571	0.0618	0.0610	0.0686
RMSE Train	0.2389	0.2486	0.2469	0.2618
R^2 Test	0.9022	0.9026	0.9052	0.9008
MSE Test	0.0978	0.0974	0.0948	0.0992
RMSE Test	0.3127	0.3121	0.3079	0.3149

R^2 observations after alpha is doubled:

- Ridge Model
  - R^2 score for train data is decreased
  - R^2 score for test data as negligible increase
- Lasso Model
  - R^2 score for train data as negligible decrease
  - R^2 score for test data as negligible decrease

Below are the Top 10 Predictor variables after alpha is doubled

- Ridge Model

features	coefficient
OverallQual_9	0.432707
OverallQual_10	0.374826
Neighborhood_StoneBr	0.297627
Neighborhood_Crawfor	0.254836
BsmtExposure_Gd	0.209994
GrLivArea	0.209089
Functional_Typ	0.205748
OverallCond_9	0.198385
Exterior1st_BrkFace	0.186421
SaleType_New	0.159200

New feature is added after alpha double and order of the features and coeff. Changed

- Lasso Model

features	coefficient
OverallQual_10	1.117244
OverallQual_9	0.935879
GrLivArea	0.380450
Neighborhood_Crawfor	0.341596
SaleType_New	0.322407
Neighborhood_StoneBr	0.303214
OverallQual_8	0.266360
Functional_Typ	0.233984
Exterior1st_BrkFace	0.223832
BsmtExposure_Gd	0.164909

New feature is added after alpha double and order of the features and coeff. Changed

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:

- Ridge model optimal lambda $\lambda = 3.5$
- Lasso Model optimal lambda $\lambda = 0.001$
- Ridge Model Train R-Square: 0.94 and Test R-Square: 0.90
- Lasso Model Train R-Square: 0.94 and Test R-Square: 0.90

Both Ridge and Lasso perform almost equally, we are going to **consider Lasso** for following reasons

1. help us by performing feature elimination
2. Model generated by lasso are generally easier to interpret
3. Lasso shrinks model coefficients to 0.

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 5 most important predictor variables are:

1. OverallQual_10
2. OverallQual_9
3. Neighborhood_StoneBr
4. GrLivArea
5. Neighborhood_Crawfor

After dropping these 5 variables, below are top 5 predictor variables

features	coefficient
OverallCond_9	0.343808
2ndFlrSF	0.334168
Neighborhood_NoRidge	0.263747
1stFlrSF	0.259503
Functional_Typ	0.228927

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:

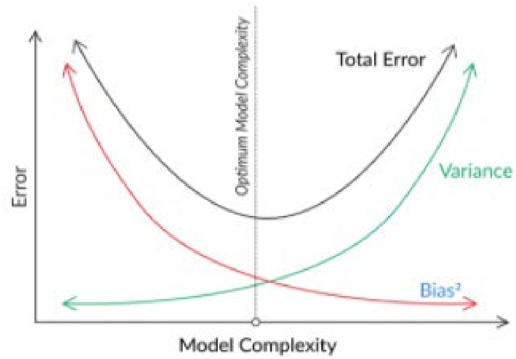
We need to ensure models neither underfitting nor overfitting.

Overfitting models perform well on training data and may not on test data. Hence, will have very low bias; but since it does not perform well with unseen data, it will show high variance.

Bias in a model is high when it does not perform well on the training data itself, and variance is high when the model does not perform well on the test data.

There is a trade-off between bias and variance with respect to model complexity, a simple model would usually have high bias and low variance, whereas a complex model would have low bias and high variance. In either case, the total error would be high.

What we need is lowest total error, i.e., low bias and low variance, such that the model identifies all the patterns that it should and is also able to perform well with unseen data.



We need to manage model complexity: It should neither be too high, which would lead to overfitting, nor too low, which would lead to a model with high bias (a biased model) that does not even identify necessary patterns in the data.

Also, the model coefficients that we obtain from an ordinary-least-squares (OLS) model can be quite unreliable if among all the predictors that we used to build our model, only a few are related significantly to the response variable.

Regularization helps with managing model complexity by essentially shrinking the model coefficient estimates towards 0. This discourages the model from becoming too complex, thus avoiding the risk of overfitting. Essentially, with regularization, we compromise by allowing a little bias for a significant gain in variance.

Lasso and Ridge adding the penalty term in the cost function which helps suppress or shrink the magnitude of the model coefficients towards 0. This discourages the creation of a more complex model, thereby preventing the risk of overfitting.