Assignment part II

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?

Ans:
Ridge and Lasso regression are some of the simple techniques to reduce model complexity and prevent over-fitting which may result from simple linear regression.

In Ridge regression, the cost function is altered by adding a penalty equivalent to square of the magnitude of the coefficients.

Alpha is the penalty term that denotes the amount of shrinking that will be implemented in the equation.

If Alpha is close to zero, the Ridge term itself is very small and thus the final error is based on RSS alone. If alpha is too large, the impact of shrinking grows and the coefficients B1, B2… Bn tends to zero.

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}\beta_j^2=\mathrm{RSS}+\lambda\sum_{j=1}^{p}\beta_j^2,$$

For lasso regression, the alpha value is 1. The output is the best cross-validated lambda, which comes out to be 0.001. once we have the optimal lambda value, we, train the lasso model.
Lasso regression takes the magnitude of the coefficients , ridge regression takes square. Ridge regression is also referred to as L2 regression.
Lasso regression can be used for automatic feature selection, as the geometry of its constrained region allows coefficient values to inert to zero.

An alpha value of zero in either ridge or lasso model will have results similar to the regression model.

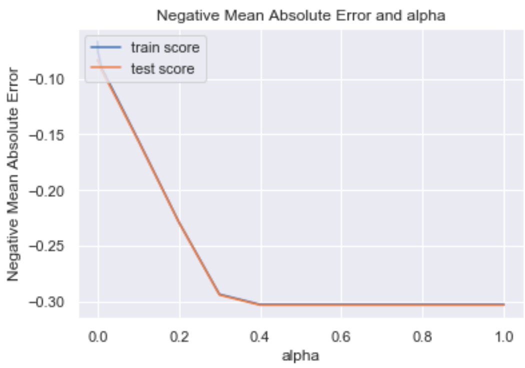
The larger the alpha value, the more aggressive the penalisation. We can say, lasso regression (L1) does both variable selection and parameter shrinkage, whereas Ridge regression only does parameter shrinkage and end up including all the coefficients in the model. In presence of correlated variables, ridge regression might be the preferred choice. Also, ridge regression works best in situations where the least square estimates have higher variance.

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?

Ans:

We applied ridge and lasso regression while predicting the housing price.

The fig., below is the plotting mean test and train scores with alpha.



From the above graph we can see that negative Mean Absolute Error is quite low at alpha=0.4 and stabilises thereafter, but we will choose a low value of alpha to balance the trade-off between Bias-variance and get the coefficients of smallest of features.

At alpha=0.01, and applying lasso even the smallest of negative coefficients that have some predictive power towards ‘SalePrice’ have been generated. The advantage of this technique is lasso brings the coefficients of significant features to zero.

Variable	Coeff	
0	constant	12,003
13	GrLivArea	0,125
4	OverallQual	0,112
5	OverallCond	0,050
9	TotalBsmtSF	0,042
7	BsmtFinSF1	0,035
21	GarageArea	0,034
20	Fireplaces	0,024
3	LotArea	0,015
2	LotFrontage	0,014

14	BsmtFullBath	0,010
22	WoodDeckSF	0,010
26	ScreenPorch	0,005
173	KitchenQual_TA	-0,007
1	MSSubClass	-0,007
19	KitchenAbvGr	-0,008
28	PropAge	-0,095

These are the 16 variables obtained from Lasso Regression that can be concluded to have the strong effect on the SalePrice.

Applying Ridge Regression and plotting mean test and train scores with alpha

Since the negative mean absolute error stabilises at alpha =2, we will choose this for further analysis. After applying RSME we get the value as RMSE : 0.11485785595060997

It is visible that the model performance is better than Lasso. The train and test scores are matching well.

Though the model performance by Ridge Regression was better in terms of R2 values of Train and Test, it is better to use lasso, since it brings and assigns a zero value to insignificant features, enabling us to choose the predictive variables.

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?

Ans:
The most important predictor variables that will be excluded are:
1. GrLivArea
2. OverallQual
3. OverallCond
4. TotalBsmtSF
5. GarageArea

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?

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

Bias: Bias is error in model, when the model is weak to learn from the data.High bias means model is unable to learn details in the data. Model performs poor on training and testing data.

Variance : Variance is error in model, when model tries to over learn from the data. High variance mean model performs well on training data as it has very well trained on this of data but performs very poor on testing data as it was unseen data for the model.
It is important to have balance in Bias and Variance to avoid overfitting and under-fitting of data.