Assignment Part - 2

Q1. *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. The optimal value of Alpha for ridge and lasso as mentioned below:

- Ridge Alpha value = 2
- Lasso Alpha Value = 0.0005

- The KPIS before doubles the alpha value of Ridge and Lasso

KPI	Ridge	Lasso
R2 Train	0.835479	0.828071
R2 Test	0.837494	0.831544
RMSE Train	0.155275	0.158732
RMSE Test	0.175248	0.178428

- The KPIS after doubles the alpha value of Ridge and Lasso , the accuracy decreased for both the models

KPI	Ridge	Lasso
R2 Train	0.828237	0.813307
R2 Test	0.833552	0.826616
RMSE Train	0.158655	0.165407
RMSE Test	0.177361	0.181018

- I have sorted the variables with Lasso as lasso was giving us good result hence the predictors before change the alpha value was

By – Akhilesh Tiwari

Variable	Ridge_coeff	Lasso_coeff
TotalCarpetArea	0.689336	7.135149e-01
ExterQual_Fa	-0.307699	-3.264331e-01
Exterior1st_BrkComm	-0.269021	-3.002200e-01
ExterQual_TA	-0.301899	-2.990194e-01
Condition2_PosN	-0.270998	-2.521616e-01

- After doubling the alpha value, now the predictors are as follows:

Variable	Ridge_coeff	Lasso_coeff
TotalCarpetArea	0.669446	7.198401e-01
ExterQual_TA	-0.275174	-2.451383e-01
ExterQual_Fa	-0.234726	-2.024648e-01
BsmtQual_NA	0.051341	1.706123e-01
SaleType_New	0.164904	1.471327e-01

Q2. 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. The optimal value of Alpha for ridge and lasso as mentioned below:

- Ridge Alpha value = 2
- Lasso Alpha Value = 0.0005

The R^2 of the Lasso is better for training as well as testing data set hence we have chosen Lasso is best estimator for this problem, below are the figures for your reference

KPI	Ridge	Lasso
R2 Train	0.835479	0.828071
R2 Test	0.837494	0.831544
RMSE Train	0.155275	0.158732
RMSE Test	0.175248	0.178428

Q3. 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?

Ans. The top five predictors which predicted the House price were as follows:

Variable	Ridge_coeff	Lasso_coeff
TotalCarpetArea	0.689336	7.135149e-01
ExterQual_Fa	-0.307699	-3.264331e-01
Exterior1st_BrkComm	-0.269021	-3.002200e-01
ExterQual_TA	-0.301899	-2.990194e-01
Condition2_PosN	-0.270998	-2.521616e-01

After removing the above variables from the data, we got below top five predictors to predict the house price

Variable	Ridge_coeff	Lasso_coeff
GrAreaPerRmsAbvGrd	0.707182	7.396550e-01
RoofMatl_WdShngl	0.374376	4.078553e-01
SaleType_New	0.328722	3.275454e-01
Functional_Maj2	-0.304966	-3.097402e-01
BsmtExposure_Gd	0.194797	1.900290e-01

Q4. 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. To avoid being impacted by outliers in the training data, a model must be made robust and generalizable. The model should also be generalizable, meaning that the test accuracy does not fall below the training score. For datasets other than the ones used during training, the model should be accurate. Outliers should not receive excessive weighting in order for the model's accuracy to be high. Outlier analysis must be performed to guarantee that this is not the case, and only those outliers that are relevant to the dataset must be kept. Those outliers in the dataset that don't make sense to keep must be deleted. This would assist improve the accuracy of the model's predictions. The usage of confidence intervals is possible (typically 3-5 standard deviations). This would help to standardized the model's predictions. It is impossible to trust a model for predictive analysis if it is not robust.

By – Akhilesh Tiwari