## Advanced Regression 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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

## Answer:

## Ridge Regression Analysis:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Alpha = 7.0 | | | | |
|  | Mean Absolute Error | Mean Squared Error | Root Mean Square Error | R2 Score |
| Train Data | 0.0705 | 0.0104 | 0.1019 | 0.9187 |
| Test Data | 0.0812 | 0.0139 | 0.1181 | 0.8882 |
| Alpha=14.0 | | | | |
|  | Mean Absolute Error | Mean Squared Error | Root Mean Square Error | R2 Score |
| Train Data | 0.07139 | 0.0106 | 0.1032 | 0.9167 |
| Test Data | 0.0822 | 0.01411 | 0.1188 | 0.8868 |

|  |  |
| --- | --- |
| Alpha=7.0 | Alpha=14.0 |
|  |  |

* Ridge Regression's ideal alpha value is 7.0. So, if, we doubled the alpha then it will become 14.
* There is slight decrease in R2 score in Test for alpha=14.
* There is change in the position of co-efficient for Alpha=14.

## Lasso Regression Analysis:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Alpha = 0.0004 | | | | |
|  | Mean Absolute Error | Mean Squared Error | Root Mean Square Error | R2 Score |
| Train Data | 0.0717 | 0.0106 | 0.1030 | 0.9171 |
| Test Data | 0.0824 | 0.0142 | 0.1192 | 0.8860 |
| Alpha=0.0008 | | | | |
|  | Mean Absolute Error | Mean Squared Error | Root Mean Square Error | R2 Score |
| Train Data | 0.0732 | 0.0110 | 0.1053 | 0.9133 |
| Test Data | 0.0835 | 0.0144 | 0.1201 | 0.8842 |

|  |  |
| --- | --- |
| Alpha= 0.0004 | Alpha=0.0008 |
|  |  |

* lasso Regression's ideal alpha value is 0.0004. So, if, we doubled the alpha then it will become 0.0008.
* There is slight decrease in R2 score in Test for alpha=0.0008.
* There is change in the position of co-efficient for Alpha=0.0008.

## 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 regression does not perform as well as Lasso regression, so we will apply it instead. It also eliminates other features by bringing the coefficients to zero, making it less complex than the Ridge model.

Lasso helps with feature reduction by setting redundant variable coefficients to 0. Lasso indirectly selects variables by reducing the coefficients of redundant variables to 0. -

Additionally, it is more tolerant of outliers.

On the other hand, ridge regression reduces the coefficients to arbitrarily low values, but not to zero.

When there is high multi-collinearity, or high correlation, between various features, Lasso Regression perform better.

## Question-3:

## 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?

## Answer:

Below are the five predictors that will have influence on the sales-

|  |  |
| --- | --- |
| Variables | Co-efficient |
| MSZoning\_RL | 0.2965 |
| MSZoning\_RH | 0.2959 |
| MSZoning\_RM | 0.2694 |
| MSZoning\_FV | 0.2615 |
| 1stFlrSF | 0.1225 |

Note: We have not removed variables based on VIF like we do it in actual prediction.

## 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:

When data variance has little to no impact on a model's performance, the model is robust. A generalizable model can react effectively to unexpectedly added data that originates from the same distribution as the model's original data.

To make sure a model is resilient and generalizable, we must be careful not to overfit it. An overfitting model has a very high variance and is exceedingly sensitive to even little changes in the data. A model of this type will be able to identify every pattern in training data, but it won't pick up on unnoticed patterns in test data.

Keep the model simple—but not overly simple, as it would be pointless.

To create a simpler model, regularization might be used. Making a model simple also results in the Bias-Variance Trade-off. Thus, we must make balance between bias and variance.

Thus, by maintaining the balance between bias and variance, which reduces the overall error as illustrated in the graph below, the accuracy of the model may be maintained.

