**Question-1**

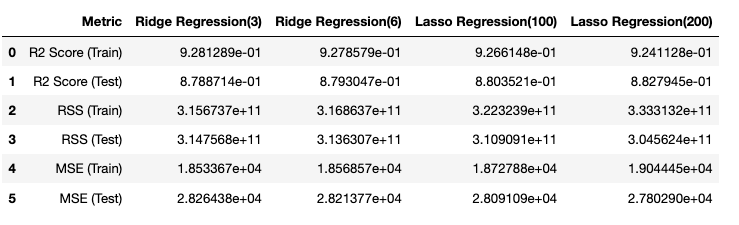
What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer:

The optimal value of alpha for ridge and lasso regression are

|  |  |
| --- | --- |
| Regression | Optimal Value of Alpha |
| Ridge | 3 |
| Lasso | 100 |

When I double the value of alpha for both ridge and lasso regression. Below are the readings:



Most important predictors after change for ridge regression

|  |  |
| --- | --- |
| Ridge (α=3) | Ridge (α=6) |
|  |  |

Most important predictors after change for lasso regression

|  |  |
| --- | --- |
| Lasso (α=100) | Lasso (α=200) |
|  |  |

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?

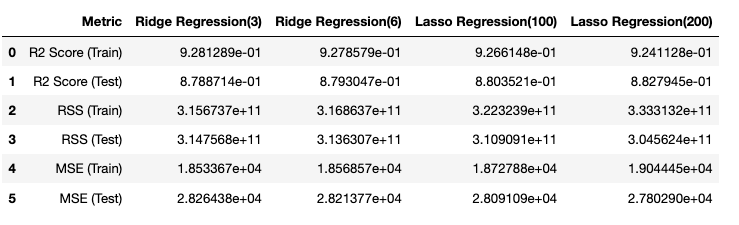
Answer:

The feature selected for Ridge and Lasso regressions are

|  |  |  |
| --- | --- | --- |
| Ridge(α=3/6) | Lasso(α=100) | Lasso(α=200) |
| 124 | 108 | 99 |

I will be selecting Lasso Regression for alpha value of 200.

1. It has minimal (99) variables for interpretation. Since Lasso shrinks some of the variable coefficients to 0 thus performing variable selection
2. R2Score, RSS and MSE are better for lasso.



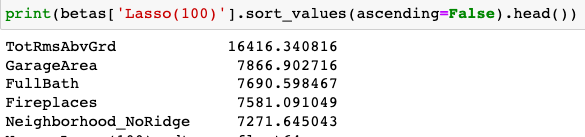
**Question-3**

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 five most important predictor variables for Lasso Regression (α=100 – optimal value) are

1. TotRmsAbvGrd: Total rooms above grade
2. GarageArea: garage area
3. FullBath: Full bathrooms above grade
4. Fireplaces: Number of fire places
5. Neighborhood\_NoRidge: NorthRidge

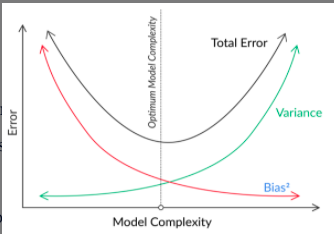


**Question-4**

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 can say that a model is robust and generalized when both Bias and Variance is low. In this case the model performs well in both test and trained data.



When Bais is low and variance is high that means model is too complex and may leads to overfitting. This kind of model performs well on train data but not on test data.

When Bais is high and Variance is low that means model is too simple and may leads to underfitting. This kind of model neither perform well on train data nor perform well on test data.

Regularization helps with managing the model complexity by essentially shrinking the model coefficients estimates towards 0. This discourage the model being too complex and thus avoiding the risk of overfitting.