

EAS 508 Homework – 2

Name: Venkata Satya Surya Sai Vineet Atyam

UB Person number: 50419767

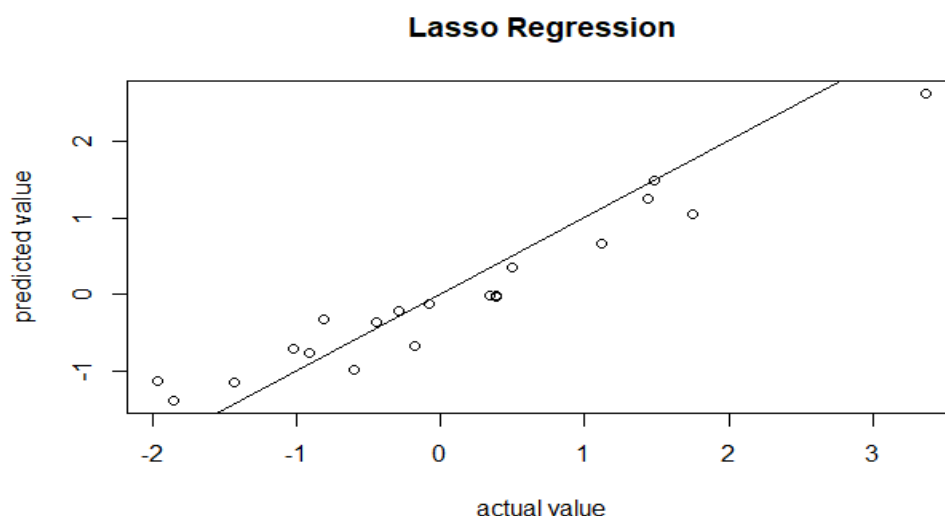
UBIT Name: vatyam

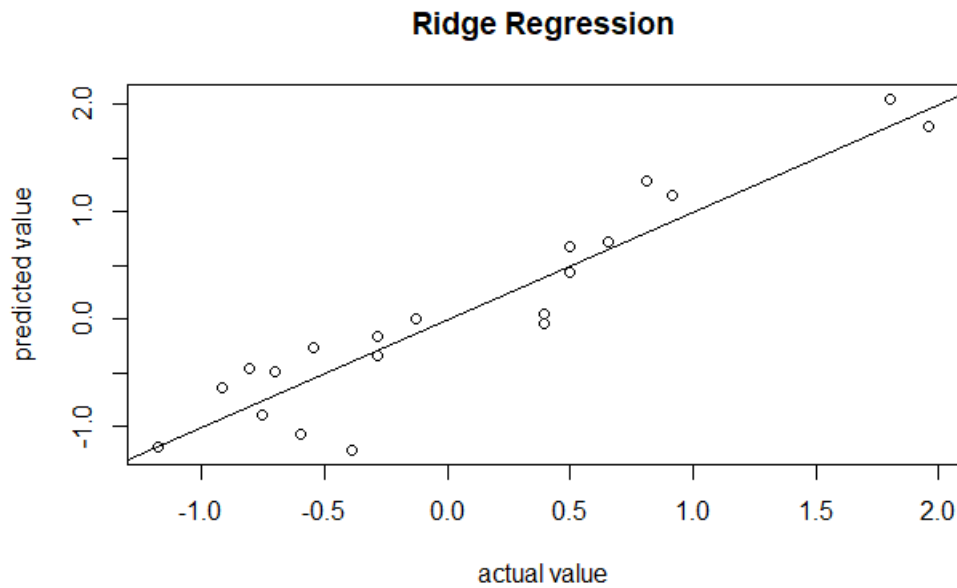
Model	RMSE_Train	RMSE_Test	R2_Train	R2_Test
Multiple Regression	0.4063055	1.437126	0.828552	-0.1559262
Ridge Regression	0.410898	0.3550886	0.8507212	0.8481744
Lasso Regression	0.4100312	0.3562787	0.8780346	0.9100231
Support Vector Regression	0.2252534	1.516017	0.7609582	-0.4697954
Gauss Process	0.3716442	0.3690241	0.8609582	0.7329183

1.) Of all the models learnt till now, for the given homework dataset, I feel the best model to work is the Lasso Regression and then Ridge Regression as they have larger R^2 and lower RMSE values.

The reason for choosing these models is that:

- They penalize the usage of more features and hence prevent over fitting, which can be seen with the above R^2 values, which are better for both Training and Test sets compared to the other regression models.
- These models have both better accuracy and also provide more robustness in predicting the unknown values and hence can work as real-life models.
- Since these models are linear model, they provide better interpretability with respect to the importance of each feature in the dataset as compared to non-linear models.





2.) Reasons for why it is better than other models used:

- **Better than Multiple Linear Regression:**

It is better than multiple linear regression because in the provided dataset, with the given data. There is huge overfitting present and hence there is a no use for the model in real conditions as it would not be able to predict them accurately. Lasso prevents overfitting and hence solves the problem with multiple regression.

- **Better than Gauss Process Regression:**

Although the accuracy metrics for gauss model also looks good and can be used for future predictions, it is avoided as it is a non-linear model, it provides less interpretability as compared to the linear model.

- **Better than Support Vector Regression:**

The SVR model is not suitable for the given dataset as it do not provide acceptable accuracy and error values and must be avoided in such cases. The SVR is not able to find a hyperplane with all the features present, but if features were reduced, the resultant model may have better predictability.