

## Advance Regression-Subjective Question

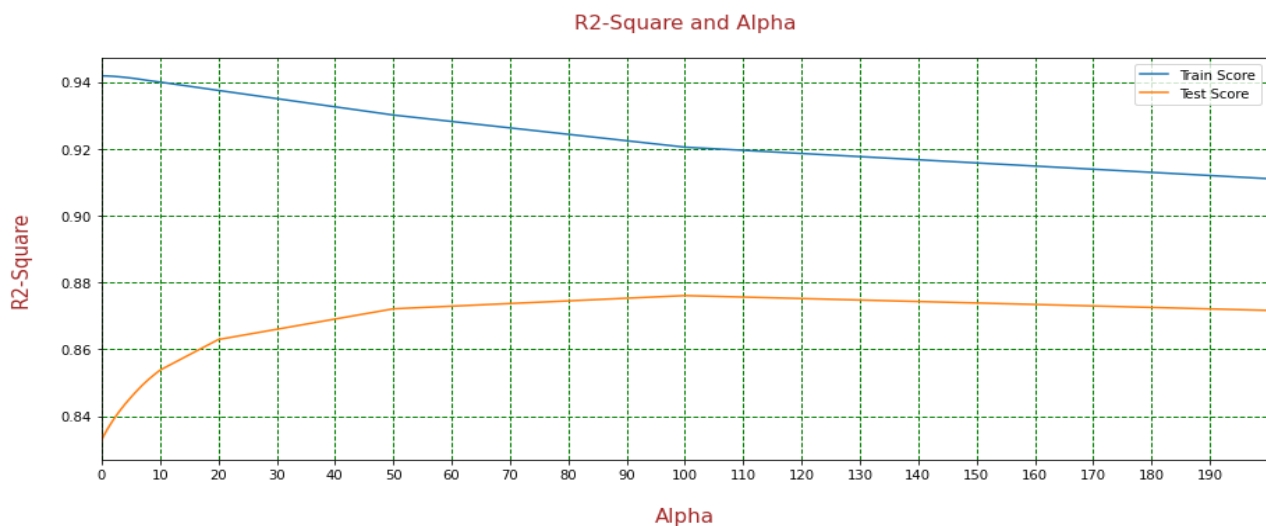
**Name:** Pieush Vyas

**Q.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?

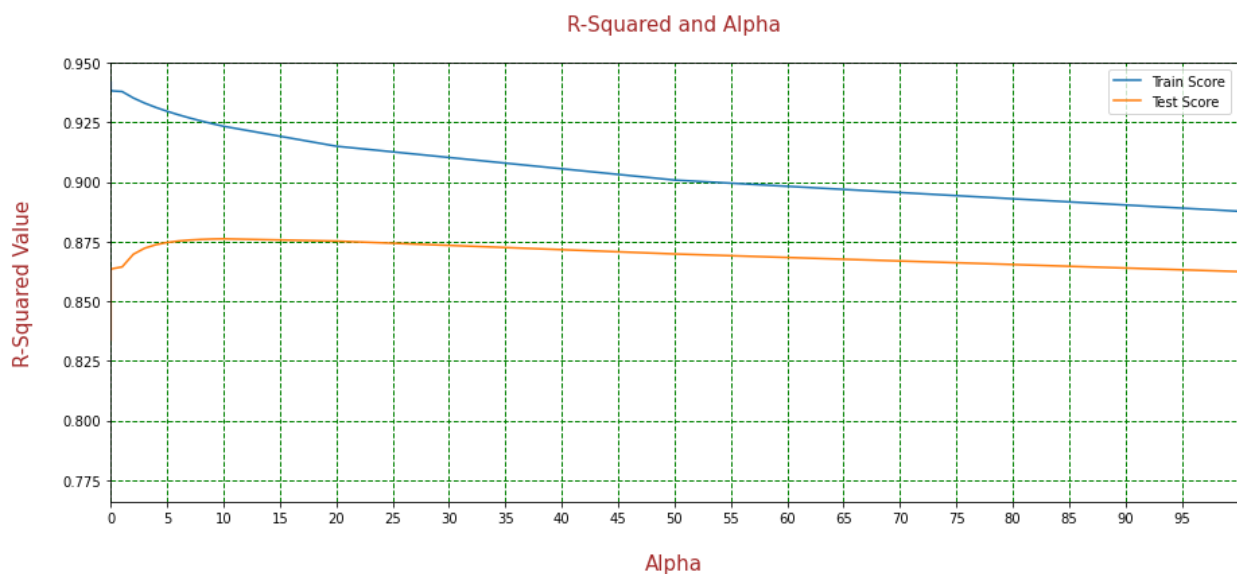
Ans.: a) Optimal value for Ridge and Lasso is 10 and 100 respectively.

b) After doubling the value of Hyper Parameter or Alpha for Ridge and Lasso the R-squared value for Train and Test data has been changed and reduced a little, as well as there is a change in top 5 features and their score for the Ridge model but there is no such change in Lasso model.

c) The most important predictor variable is 'GrLivArea' which has highest score.



**Fig.1. Lasso (Alpha=100)**



**Fig.2. Ridge (Alpha=10)**

**Q.2:** You have determined the optimal value of lambda for Ridge and Lasso during the assignment. Now, which one will you choose to apply and why?

**Ans.:** I will choose 10 for Lambda (Alpha) derived for Ridge model as at this value R-Squared value of Train and Test data is highest and the Ridge is faster than Lasso for feature selection specially for the large data.

**Q.3:** After building the model, you realized that the five 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 five important predictor variables now?

**Ans.:** Previously the top 5 predictors were

```
{'GrLivArea'- 18522.457,  
'MSZoning_FV' - 16726.251,  
'Neighborhood_Crawfor' - 14582.784,  
'YearBuilt' - 11716.072,  
'OverallQual' - 11290.349}.
```

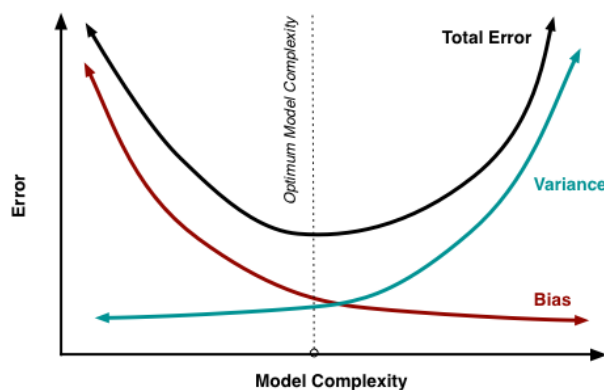
**After removal of top 5 variables from the list new top 5 variables are**

```
{'TotalBsmtSF' - 19508.675,  
'2ndFlrSF' - 17372.919,  
'Neighborhood_OldTown' - (-) 15167.369,  
'Neighborhood_Edwards' - (-) 14519.661,  
'Neighborhood_Gilbert' - (-) 12997.745}
```

**Q.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?

**Ans.:** Model is considered robust and generalizable...

- When the model is Stable and Simple, means having a tradeoff between Bias and Variance or with a reasonable level of variance and accuracy.
- When the model has not much affected by the change in training data and do not overfit the dataset.
- Accuracy should be almost same or there should be no much discrimination for the training and test data. Such model may perform equally well for both the data (Training and Test data).
- There should be very less Error between Train and Test data.



**Fig.3. Bias-Variance Trade-Off (Ref. UpGrade Notes)**