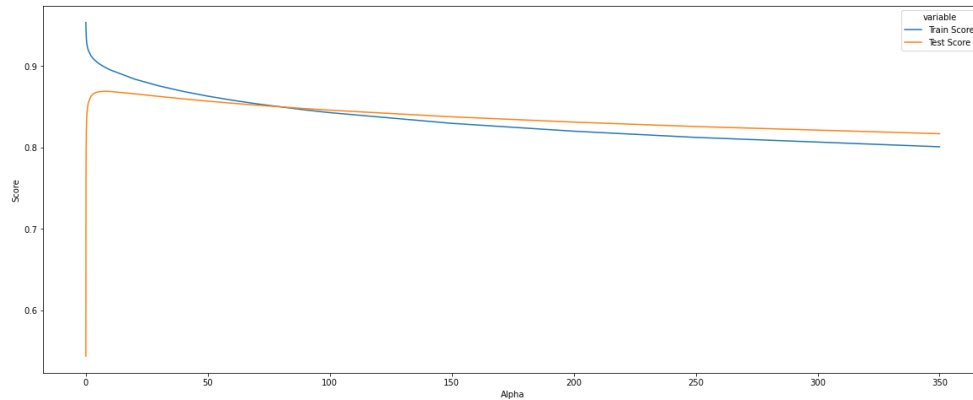


Question 1

Optimal Value for Ridge is 10

Reason :

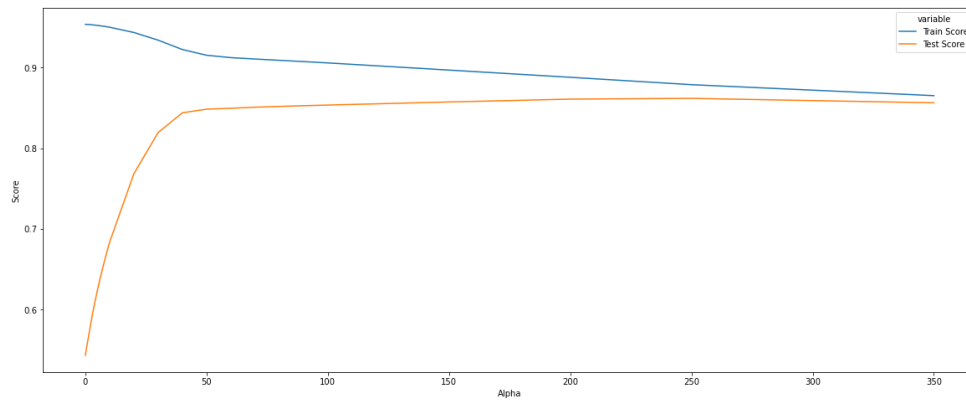
As observed in the above plot train and test scores starts to stabilize from 10



Optimal Value for Lasso is 40

Reason :

As observed in the above plot train and test scores starts to stabilize from 40



Doubling Alpha for Ridge Regression

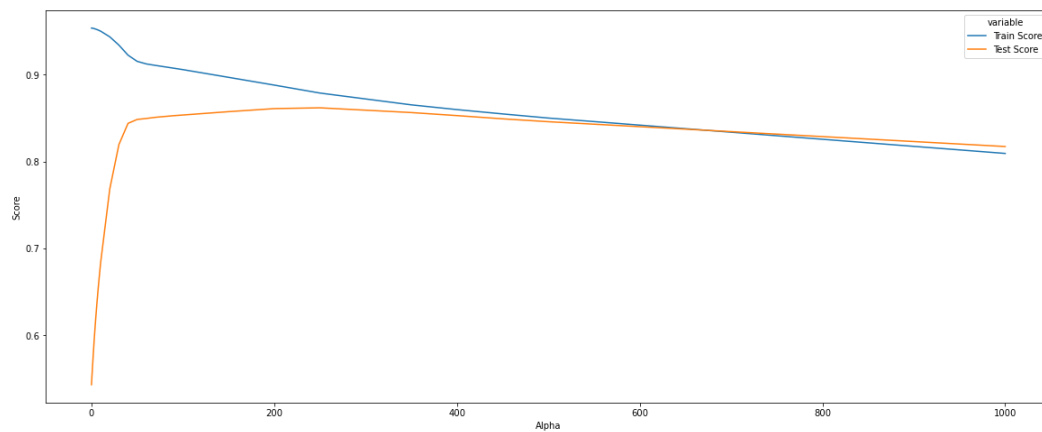
Train accuracy goes down by 1% (89%→88%) and test accuracy does remains same.

Similar trend of decreasing train accuracy follows as we increase the alpha value due increase in regularization an increasing bias as the result.

Doubling Alpha for Lasso Regression

Train accuracy goes down by 2% (92% -> 90%) increasing test accuracy by 1% (84% -> 85%)

Test and Train accuracy scores keep coming closer to each other i.e. decreasing train and slightly increasing test score until they dissect each other at (~400) then follow the same trend with increasing values of alpha as shown in the below plot



Question 2

I would choose Lasso over Ridge

- Ridge regression can't zero out coefficients so, you either end up including all the coefficients in the model. As this model has close to 288 features including dummy variables, including all these features in the model would make it fairly complex
- Lasso does both parameter shrinkage and variable selection automatically
- As Lasso does feature elimination it would decrease the model complexity.

Either way both the models are good with just minor difference

Question 3

	Lasso_Pred	Lasso_Rank	Ridge_Pred	Ridge_Rank
8	RoofMatl_WdShngl	129	BsmtExposure_Gd	246
7	RoofMatl_WdShake	140	OverallQual_Very Good	250
0	FullBath	142	FullBath	248
9	BsmtQual_NA	151	BsmtFinType1_NA	262
2	Neighborhood_NoRidge	153	Neighborhood_Crawfor	259
5	OverallQual_Very Excellent	159	OverallQual_Excellent	269
3	Condition2_PosN	160	Neighborhood_NoRidge	267
1	Neighborhood_Crawfor	172	GarageCars	271
4	OverallQual_Excellent	173	Neighborhood_NridgHt	270
6	OverallQual_Very Good	174	OverallQual_Very Excellent	268

Top predictors extracted by performing RFE on Lasso are

- Neighborhood_NoRidge
- Condition2_PosN
- OverallQual_Excellent
- OverallQual_Very Excellent
- OverallQual_Very Good

Top5 Predictors after removing these features are

- RoofMatl_WdShngl
- BsmtQual_Fa
- BsmtQual_Gd
- BsmtQual_TA
- BsmtCond_NA

Top predictors extracted by performing RFE on Ridge are

- FullBath
- Neighborhood_NoRidge
- OverallQual_Excellent
- OverallQual_Very Excellent
- OverallQual_Very Good

Top5 Predictors after removing these features are

- BsmtQual_Fa
- BsmtQual_Gd
- BsmtQual_TA
- BsmtFinType1_NA
- KitchenQual_TA

Question 4

Model has to be made robust and generalizable due to the bias and variance Tradeoff exists in the dataset. High Variance suggests outliers in the data, also checking if there exists multicollinearity in the features selected. Model has to be generalized so that there should not be a significant difference between test and train scores. The model has to be accurate on the test set not just on train data.

When model is accurate on Train data and not on test data it suggests the model is overfitted and using regularization would decrease variance in the data which handles the outliers effecting the model.

Removing features which multicollinearity and identifying important predictors which increase the accuracy of the model also boosts the performance of the model on test data.