Problem Statement - Part II-Subjective Questions

Assignment Part-II

Submitted by Sreekumar R

(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?

A: Alpha is otherwise called the hyperparameter which must be tuned to arrive at an optimum value to reduce overfitting. Optimum value of alpha is the one which minimizes the cost functions in Ridge and Lasso models. Optimal values of alpha observed from gridsearch algorithm and model fitting in the case of the assignment are:

Ridge Model: Alpha = 4.0

Lasso Model: Alpha = 0.0002

The below table shows the effect on the metrics on doubling the alpha values for Ridge and Lasso Models.

Ridge Model	Alpha = 4	Alpha = 8	Difference
r2 for train data	0.933654276	0.925313299	-0.008340977
r2 for test data	0.896475443	0.894517844	-0.001957598
mse for train data	0.001146087	0.001290173	0.000144086
mse for test data	0.001638144	0.001669121	0.000031
mae for train data	0.022762887	0.024193985	0.001431098
mae for test data	0.028542941	0.028174885	-0.000368055

Lasso Model	Alpha = 0.0002	Alpha = 0.0004	Difference
r2 for train data	0.928145925	0.917830305	-0.01031562
r2 for test data	0.895461936	0.889499244	-0.005962692
mse for train data	0.00124124	0.001419437	0.000178197
mse for test data	0.001654182	0.001748534	0.000094
mae for train data	0.023720012	0.025488002	0.00176799
mae for test data	0.028745686	0.028638103	-0.000107582

Doubling the hyperparameter makes the model simpler. It can be observed that the r2 score decreases marginally for both test data and train data for these models arrived at in the assignment. MSE and MAE remains almost the same though theoretically the error values increase. R2 score also is getting reduced. Increase in error values and reduction in r2 score though marginal implies less optimal model.

The most influential predictor variables (considered first 5nos here) remains the same after doubling of alpha though the values of the coefficients reduced marginally. The following table

summarizes the first five nos most influential variables and the corresponding coefficients after doubling the alpha values.

Influential variables in case of Lasso Model		
Variable name	coefficient after doubling alpha	Coefficients before doubling alpha
TotalBsmtSF	0.044090005	0.051212617
BsmtFinSF1	0.055935946	0.056762251
Neighborhood_StoneBr	0.067260207	0.081240342
OverallQual_9	0.104594233	0.097439143
GrLivArea	0.288853162	0.293785232

Influential variables in case of Ridge Model		
Variable name	Coefficients after doubling alpha	Coefficients before doubling alpha
BsmtFinSF1	0.053512718	0.061195205
TotalBsmtSF	0.05372473	0.063333841
OverallQual_9	0.057724596	0.064576906
GrLivArea	0.062998721	0.076976762
1stFlrSF	0.069328185	0.081107598

(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?

A: Lambda is the hypertuning parameter in Ridge and Lasso regression models. This has to be tuned to arrive at an optimum value to reduce overfitting. Optimal values of alpha observed from gridsearch algorithm and model fitting in the assignment are:

Ridge Model: Alpha = 4.0

Lasso Model: Alpha = 0.0002

The below table shows the metrics for Ridge and Lasso Models.

	Ridge Model Alpha = 4	Lasso Model Alpha = 0.0002
r2 for train data	0.933654276	0.928145925
r2 for test data	0.896475443	0.895461936
mse for train data	0.001146087	0.00124124
mse for test data	0.001638144	0.001654182
mae for train data	0.022762887	0.023720012
mae for test data	0.028542941	0.028745686

It can be seen from the above table that the r2 score is marginally higher for the Ridge model for both training and test data sets. In addition, the mean square error and the mean absolute errors are also lesser in the case of Ridge model. Therefore, Ridge model will be the

chosen one when going ONLY by the metrics. But, considering model simplicity LASSO is the preferred one, explanation is given below.

It has reduced the coefficients of 135 nos of variables to zero. Lasso regression model inherently is a feature selection model when compared to ridge model. In this specific case of assignment, ridge model reduces the number of features only by 11, in comparison to lasso where 135 nos were reduced.

Total number of non-zero feature coefficients in Lasso model: 237 – 135 = 102 features

Total number of non-zero feature coefficients in Ridge model: 237 – 11 = 126 features

Looking at the complexity of the models, lasso model is much simpler, and has comparable metrics too.

So, giving an upper hand to simpler models and also considering the metrics, LASSO model is the preferred on this use case.

(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?

A: The below given table shows the list of 5 most influential variables in the case of Lasso regression model.

First two columns give the variables and corresponding coefficients before excluding the most influential variables and the second two columns give the new variables and corresponding coefficients after removal.

Informatively it is observed that the value of alpha remains the same after removal of the most influential variables. The value of alpha is still **0.0002.**

Most Influential variables: before	Corresponding Coefficients:	New Most Influential variables: after	Corresponding Coefficients:
TotalBsmtSF	0.051212617	LotArea	0.041909407
BsmtFinSF1	0.056762251	Neighborhood_Crawfor	0.0423846
Neighborhood_StoneBr	0.081240342	YearBuilt	0.064478558
OverallQual_9	0.097439143	2ndFlrSF	0.161125618
GrLivArea	0.293785232	1stFlrSF	0.284381268

(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?**

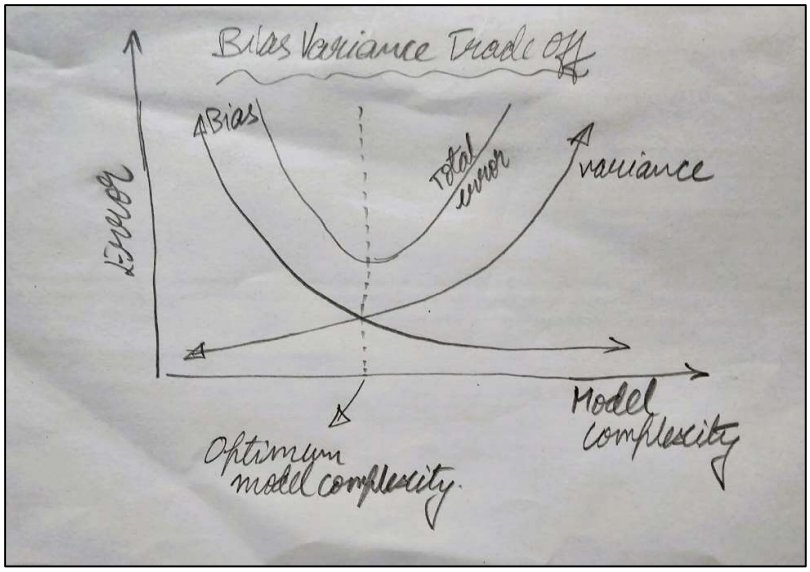
A: A model can be termed as robust and generalizable if the metrics on the training data and the test data are comparable. There should not be much difference between the training data metrics and the test data metrics rathe the difference should be acceptable. In the present assignment use case, the test data metrics are not much different from training data metrics.

If there is vast difference between the training data metrics and test data metrics, the model cannot be termed as robust and generalizable. R2 score lesser in test data than training data means there is a possibility of overfitting. Model when handling unseen data fails to some

extent. This is not the case of a robust and generalized model. Similarly, there is a case of underfitting too when there is not enough data points or features in the training data. This is also not the case of robust and generalized model. Outliers, multi-collinearity, large number of features etc can be hindrance towards arriving at a generalized and robust model. Removing outliers, removing multi-collinearity, choosing only the required number of features can all lead to a robust and generalizable model. Cross validation also helps in this. Scaling the variables can take care of outliers to some extent, so also using Ridge and Lasso models etc can also lead to a robust model.

A robust and generalized model will perform almost equally well on unseen data just as it performs on the training data. For that the model should be as simple as possible. Occam's Razor principle suggests the same, go for the simplest model possible.

A highly accurate model can be very complex and may not perform well when handling unseen data. A highly accurate model may not be robust ie the variance may be high. An optimal model is a balance between accuracy and robustness. Bias-variance trade-off is to be followed where we find a point where the total error is the least. Bias refers to the accuracy of the model in predicting unseen data. Variance refers to the flexibility. A trade-off is required as shown in the below figure.



****