Problem Statement - Part II

Name – Vibhu Dutt Raturi

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 1

Optimal Value of Ridge regression – 0.8

Optimal Value of Lasso regression – 0.0001

```
LassoModel_cv.best_estimator_
```

```
▾ Lasso
Lasso(alpha=0.0001)
```

We doubled the alpha value for Ridge and lasso let's see the changes

So, For Ridge, alpha becomes 1.6 and for Lasso 0.0002

	Metric	Ridge Regression-RFE	Lasso Regression	Double-Ridge Regression-RFE	Double-Lasso Regression
0	R2 Score (Train)	0.938974	0.942055	0.936670	0.935877
1	R2 Score (Test)	0.900078	0.905762	0.899987	0.901766
2	RSS (Train)	1.086921	1.032045	1.127950	1.142077
3	RSS (Test)	0.656015	0.618694	0.656607	0.644932
4	MSE (Train)	0.001067	0.001013	0.001107	0.001121
5	MSE (Test)	0.001501	0.001416	0.001503	0.001476
6	RMSE (Train)	0.032660	0.031825	0.033270	0.033478
7	RMSE (Test)	0.038745	0.037627	0.038763	0.038416

- Now if we observe the visual above with increase if alpha value there is a slight decrease in train r2 Score and test r2 Score in ridge regression
- Train score and test score for lasso slightly decreased
- There is increase in Residual sum of square (RSS) value for train in ridge and almost same in test
- RMSE value in ridge regression is almost the same in test data and train data
- The original lasso model had 106 features as the rest were made 0 because of feature selection whereas the updated lasso model with double alpha value has 79 features.

	Double-ridgeVariables	ridgeCoefficients	Double-lassoVariables	lassoCoefficients
0	GrLivArea	0.210426	GrLivArea	0.293686
1	OverallQual	0.156402	OverallQual	0.136093
2	TotalBsmtSF	0.144898	TotalBsmtSF	0.096712
3	MSZoning_FV	0.084339	OverallCond	0.080949
4	totalSqrFootage	0.081779	totalSqrFootage	0.072648
5	OverallCond	0.081243	PropertyAgeWhenSoldinMonths	-0.037696
6	PropertyAgeWhenSoldinMonths	-0.080630	SaleType_New	0.034523
7	MSZoning_RL	0.061806	Neighborhood_Crawfor	0.032515
8	MSZoning_RH	0.058154	GarageCars	0.030053
9	TotRmsAbvGrd	0.054484	MSSubClass_160	-0.029404

These are the most important 10 features in ridge and lasso regression with their respective coefficient values once the alpha value is doubled

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 2

We know how model complexity can be an issue while model building and how sensitive a model becomes if there is a slight change in the data as well. As per Occam's razor rule, we have learnt a simple model is better than the complex model ad is more generalised.

This answer would change for data to data according to me

For this data both the models have performed good let's compare their metrics once first

	Metric	Ridge Regression-RFE	Lasso Regression
0	R2 Score (Train)	0.938974	0.942055
1	R2 Score (Test)	0.900078	0.905762
2	RSS (Train)	1.086921	1.032045
3	RSS (Test)	0.656015	0.618694
4	MSE (Train)	0.001067	0.001013
5	MSE (Test)	0.001501	0.001416
6	RMSE (Train)	0.032660	0.031825
7	RMSE (Test)	0.038745	0.037627

The features selected by Lasso Model – 106

The features selected by RFE and then used in Ridge Model – 54

Now if we talk about metrics, we have seen both are performing equally good but yes Lasso is slightly better as its RMSE is low and has good r2 value as well.

But we also know that lasso regression has a feature selection method and hence with change in values and incoming data it will accordingly change and accordingly will have top predictors for the target variable for example when alpha value in lasso was doubled the predictors went down from 106 to 79 features so it becomes an advantage when we have many features in our dataset it is a useful method for feature selection.

When it comes to Ridge model, we used it with RFECV. Which means Recursive Feature selection cross validation method with ridge regression. Where with RFECV we got the best 54 features in the dataset and then we applied ridge for regularization of the model.

Now to conclude

Lasso model used 106 features and gave us a train accuracy of 94.2% and test accuracy of 90.5%

And RMSE train – 0.031 and RMSE test – 0.037

whereas

RFE-Ridge model used 54 features and gave us a train accuracy of 93.8% and test accuracy of 90%

And RMSE train – 0.032 and RMSE test – 0.038

So, with a difference of 52 features between Ridge and Lasso, Ridge is performing almost the same when compared to the Lasso model which makes it more generic and robust

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 3

These were the 5 most important predictors according to lasso regression:

lassoVariables	lassoCoefficients
GrLivArea	0.301773
OverallQual	0.131078
TotalBsmtSF	0.095186
OverallCond	0.086983
totalSqrFootage	0.077744

After removing these variables from the data and building the model again let’s see the result, we got from the updated lasso model

	lassoVariables	lassoCoefficients
0	TotRmsAbvGrd	0.186622
1	totalBathrooms	0.106549
2	GarageArea	0.098334
3	MSZoning_FV	0.084201
4	MSZoning_RL	0.078057
5	MSZoning_RH	0.075554
6	MSZoning_RM	0.067698
7	Neighborhood_Crawfor	0.058493
8	LotFrontage	0.056461
9	Neighborhood_StoneBr	0.052344

These are the top 10 features along with their respective coefficient values predicted by the updated lasso model after removing the top 5 predictors from the incoming data.

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 4

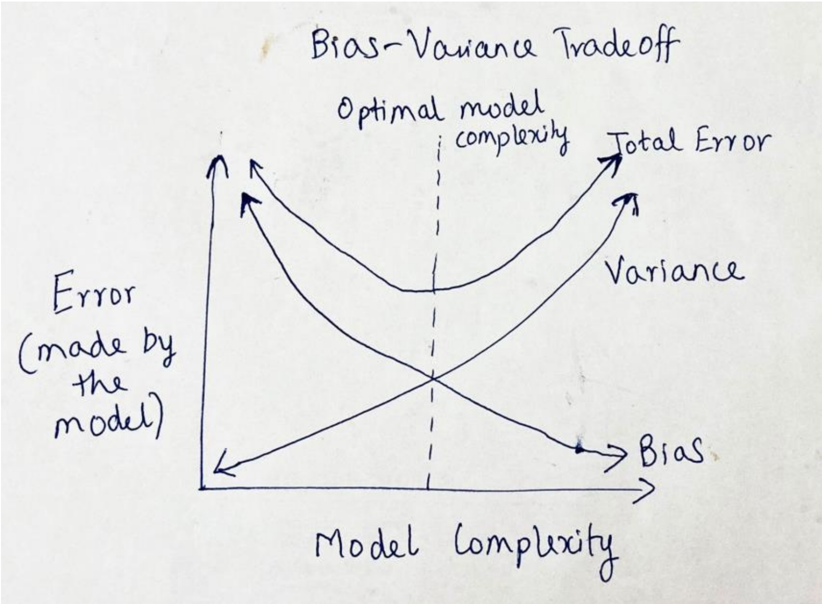
When it comes to model being robust and generalisable, Occam's razor is something which we can discuss which is essentially about making our model simpler with fewer coefficients over a complex model.

- Simpler models are usually more **generic**. If a model is complex, it will make far too many assumptions and will be pattern specific. Extra assumptions are more likely to go wrong when a model sees unseen data. Simpler models mean's keeping assumptions to minimum and likelihood of going wrong minimises hence model is more **generalised**.
- Simpler models are more Robust. What if the pattern of the data changes completely or some of the data changes. The complex model which is being trained on specific pattern will fail. Input data if changed complex model will need to be remodelled or reoriented whereas a simpler model is **Robust** as it is not specific pattern oriented.
- Simpler models do not mean that the model should be simplified to the maximum extent it means it should be simple and Robust.

Let's see how accuracy is affected

- Simpler models make more errors in the training of the model. It is possible that more generic model makes more errors than not so generic or complex models but complex models actually suffer when given unseen data.
- Simpler models require fewer training data

When it comes to accuracy of the model, model being robust and generic bias-variance trade off comes into picture. We need to establish a correct balance between bias and variance or between simplicity and complexity of a model



Underfitting model will have high bias and low variance which means model complexity is too low. The left side of the image shows what is being said.

Overfitting model will have high variance and low bias which means model complexity is too high. The right side of the image shows overfitting.

Hence, we need to balance the complexity of the model and make sure it balances out on complexity and simplicity.

Regularization is something which helps us in reducing overfitting and ensures that the model is not complex. Regularization is a part of the learning algorithm itself and it prevents the model from becoming complex. In regularization, hyperparameters are the parameters we pass on to the learning algorithm just like we had alpha/lambda used in Ridge and lasso. Ridge and Lasso are regularization techniques. Regularization shrinks the beta coefficients towards zero thus helps in reducing overfitting. Making the model robust and generic.