Advanced Regression

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?

Ans: From the experimentation:

- Optimal value of lambda for Ridge regression is 0.4
- Optimal value of lambda for Lasso regression is 0.0001

After doubling the values for ridge and lasso, we get 0.8 and 0.0002 lambdas respectively.
And following are the R2 scores –

Ridge regression:

- R2 Score (Train set) for lambda 0.4 = 0.897
- R2 Score (Test set) for lambda 0.0001 = 0.887
- R2 Score (Train set) for lambda 0.8 = 0.895
- R2 Score (Test set) for lambda 0.0002 = 0.887
-

Lasso regression:

- R2 Score (Train set) for lambda 0.4 = 0.892
- R2 Score (Test set) for lambda 0.0001 = 0.873
- R2 Score (Train set) for lambda 0.8 = 0.886
- R2 Score (Test set) for lambda 0.0002 = 0.865

When lambda values are doubled, we can see slight decrease in R2 values for Lasso model and there is no difference for Ridge model.

Variables that are most important predictors –
Ridge:

Feature	Ridge
GrLivArea	0.288356
OverallQual	0.189391
OverallCond	0.108026
MSZoning_FV	0.106682
MSZoning_RL	0.091878
MSZoning_RH	0.088834
GarageCars	0.085679
TotalBsmtSF	0.080521
Functional_Sev	-0.078458

Lasso:

Feature	Lasso
GrLivArea	0.295578
OverallQual	0.209447
OverallCond	0.106492
TotalBsmtSF	0.096517
GarageCars	0.085805
BsmtFullBath	0.051339
MSZoning_FV	0.051184
BsmtFinSF1	0.049547
KitchenAbvGr	-0.045487

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?

Ans: From the experimentation

Ridge regression:

- R2 Score (Train set) for lambda 0.4 = 0.897
- R2 Score (Test set) for lambda 0.0001 = 0.887

Lasso regression:

- R2 Score (Train set) for lambda 0.4 = 0.892
- R2 Score (Test set) for lambda 0.0001 = 0.873

The R2 score for Ridge regression is observed slightly better than Lasso regression.
Ridge regression shrinks the coefficients towards 0 but not exactly 0, which implies that all the 50 (got using RFE Method) features are included in Ridge model. When the dataset has huge number of features, it gets difficult and time consuming to use Ridge regression as it includes all the features in the model. As the number of features increases the model becomes more complex.
On the other hand, Lasso regression pushes some of the coefficients exactly to 0, which means feature selection is done. In our experiment with optimal value of lambda as 0.001, Lasso selected 31 significant features out of 50, while ridge regression contains all 50 features.
Hence, better to use Lasso regression as it does feature selection and reduces model complexity.

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?

Ans: The 5 top most variables that are significant for lambda 0.0001 in Lasso regression are –

1. GrLivArea (0.307061)
2. OverallQual (0.200335)
3. OverallCond (0.108886)
4. TotalBsmtSF (0.10485)
5. MSZoning_FV (0.084748)

In case, above features are not present in the incoming data. We will create another model using Lasso regression with lambda 0.0001 and predict the new significant features.

1. BsmtFinSF1 (0.176061)
2. FullBath (0.175214)
3. GarageCars (0.16926)
4. LotArea (0.133107)
5. BsmtUnfSF (0.122946)

R2 score changes observed are as follows –
Lasso (without feature removal)

- R2 Score (Train set) for lambda 0.4 = 0.892
- R2 Score (Test set) for lambda 0.0001 = 0.873

Lasso (top 5-feature removal)

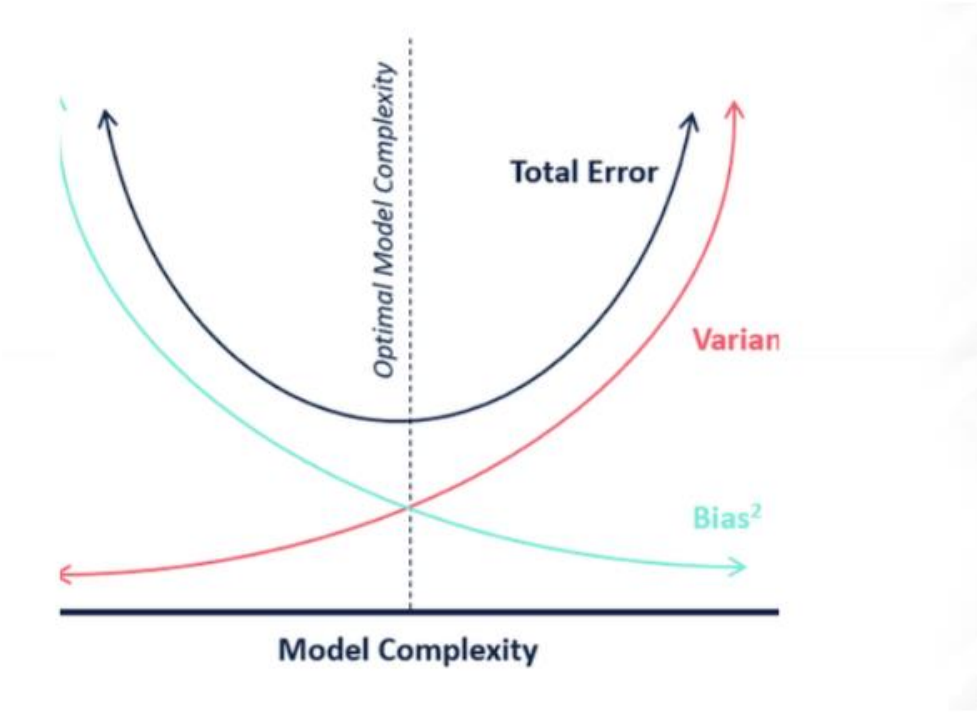
- R2 Score (Train set) for lambda 0.4 = 0.786
- R2 Score (Test set) for lambda 0.0001 = 0.665

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?

Ans: If the model works well on the train and test set, it is said to be robust i.e no underfitting and overfitting. Simple models can be easily generalizable though the bias is

reasonably compromised. A model should be as simple as necessary and not simpler than that. There is always a trade-off between bias and variance.



Bias

The bias is known as the difference between the prediction of the values by the ML model and the correct value. Being high in biasing gives a large error in training as well as testing data

Variance

The variability of model prediction for a given data point which tells us spread of our data is called the variance of the model. The model with high variance has a very complex fit to the training data and thus is not able to fit accurately on the data which it hasn't seen before.

Hence, to find balance between bias and variance we use regularisation and avoid overfitting and underfitting.