Contents

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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Ridge Regression:

Optimum alpha value	20
R-Squared (Train)	0.93
R-Squared (Test)	0.93

Coefficients for the most important predictor variables

GrLivArea	0.08
Neighborhood_Crawfor	0.07
OverallQual	0.06
Exterior1st_BrkFace	0.05
Functional_Typ	0.05
Condition1_Norm	0.04
SaleCondition_Alloca	0.04
SaleCondition_Normal	0.04
OverallCond	0.04
TotalBsmtSF	0.04

Doubling the alpha value

alpha value	40
R-Squared (Train)	0.92
R-Squared (Test)	0.93

Coefficients for the most important predictor variables

GrLivArea	0.07
OverallQual	0.06
Neighborhood_Crawfor	0.05
Functional_Typ	0.04
OverallCond	0.04
TotalBsmtSF	0.04
Exterior1st_BrkFace	0.04
Condition1_Norm	0.04
2ndFlrSF	0.03

After doubling the alpha values:

Changes in Ridge Regression metrics:

- R2 score of train set decreased from 0.93 to 0.92.
- R2 score of test set remains same at 0.93.

Lasso regression:

Optimum alpha value	0.001
R-Squared (Train)	0.92
R-Squared (Test)	0.93

Coefficients for the most important predictor variables

GrLivArea	0.101727
Neighborhood_Crawfor	0.087335
Exterior1st_BrkFace	0.07034
OverallQual	0.069625
Functional_Typ	0.05795
Condition1_Norm	0.04668
TotalBsmtSF	0.043827
OverallCond	0.041983
Neighborhood_Somerst	0.037256
Neighborhood_NridgHt	0.037215

Doubling the alpha value

alpha value	0.002
R-Squared (Train)	0.91
R-Squared (Test)	0.92

Coefficients for the most important predictor variables

GrLivArea	0.100275
OverallQual	0.080562
Neighborhood_Crawfor	0.061374
TotalBsmtSF	0.043759
OverallCond	0.041183
Condition1_Norm	0.040549
Foundation_PConc	0.03843
Functional_Typ	0.037674
Exterior1st_BrkFace	0.03736
GarageArea	0.033237

After doubling the alpha values:

Changes in Lasso Regression metrics:

- R2 score of train set decreased from 0.92 to 0.91.
- R2 score of test set decreased from at 0.93 to 0.92.

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?

- To avoid overfitting, we are using regularization techniques.
- Penalty is imposed to regularize.
- Ridge regression uses L2 on the other hand lasso regression uses L1 regularisation technique.
- In ridge regression,
    o the penalty is equal to the sum of the squares of the coefficients.
- In the Lasso regression,
    o penalty is the sum of the absolute values of the coefficients.

- Ridge shrinks the coefficients, but coefficients cannot be zero.
- Lasso shrinks the coefficients and coefficients can become zero hence it does feature selection.
- In cases where only a small number of predictor variables are significant, lasso regression tends to perform better because it's able to shrink insignificant variables completely to zero and remove them from the model.
- when many predictor variables are significant in the model and their coefficients are roughly equal then ridge regression tends to perform better because it keeps all the predictors in the model.
- In our case, both Lasso and Ridge give almost same result with respect to R2 score for test and train data for the optimum alpha, however Lasso performs feature selection.

For further information I referred
https://learn.upgrad.com/course/4622/segment/42429/247803/756832/3806923 the reason why Lasso performs feature selection through geometrical Representation of Ridge and Lasso section.

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?

Top 5 Important predictors

| GrLivArea |
|---|
| Neighborhood_Crawfor |
| Exterior1st_BrkFace |
| OverallQual |
| Functional_Typ |

After removing first 5 important predictors and remodelling below are the top 5 important predictors.

MSSubClass_2-STORY 1945 & OLDER
OverallCond
Neighborhood_StoneBr
Neighborhood_NridgHt
Condition1_Norm

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?

- The model should neither be very complex / overfit nor be very simple / underfit.
- Overfitting is a model which is highly specific to the data on which it is trained but fails to generalise to other unseen data points i.e A model is more specific to training but fails to perform well for testing data or unseen data points.
- Underfitting is a model which is very simple and fails to represent even the training data well.
- For overfitting variance is very high variance, I.e since the model memorizes for the given training set, when there is any change in the data the model will also need to change drastically. This is unstable and sensitive to changes in the training data.
- Variance refers to the degree of changes in the model itself with respect to changes in the training data.
- Bias is how accurate the model is likely to be on future data i.e test or unseen data. Extremely simple models are likely to fail in predicting complex real-world phenomena.
- With high accuracy model -> high variance low bias
- With simple model -> low variance high bias
- There should be bias-variance trade off.

  o

Ref: https://learn.upgrad.com/course/4622/segment/27474/243296/743380/3749884
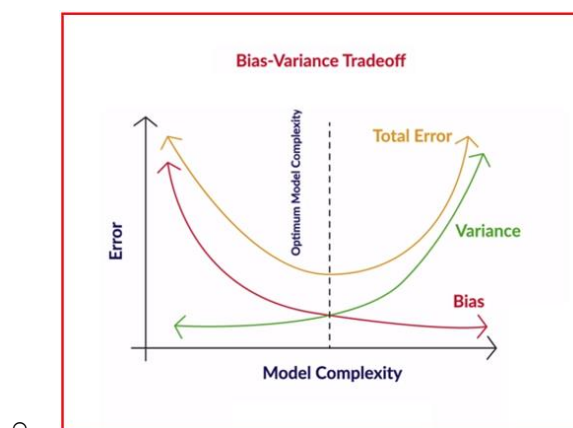
- To make our model more robust and generalizable, is to keep our model optimum to bias and variance. To minimize the overall error. Regularization Ridge/ Lasso regularization help to reduce the variance.

Note: In our model, Ridge and Lasso provided good R2 score for both training and test data