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:

Ridge Regression: The model built on ridge regression at an optimal value of alpha has:

- Optimal value (alpha) = 3
- R2 value of train data is =0.92
- R2 Value of test data is = 0.90

Top features with coefficients:

Feature	Co-efficient
Total SF	0.39
TotalBsmtSF	0.23
OverallQual_9	0.18
GarageArea	0.16
LotArea	0.16

By adding the hyperparameter (alpha) which is lambda times the residual sum of squares we can make the fit less complex. As lambda increases the bias is unchanged but the variance drops. The drawback is it includes all the variables but try to shrink the coefficients to zero. So, the larger the value of alpha the coefficients of the features will tend to shrink to zero.

After doubling the alpha value:

- (alpha) = 6
- R2 value of train data is =0.92
- R2 Value of test data is = 0.89

Top predictor variables after the alpha value is doubled:

Feature	Co-efficient
Total SF	0.30
TotalBsmtSF	0.19
OverallQual_9	0.16
BsmtFinSF1	0.15
GarageArea	0.14

The predictor variables are almost same but the coefficients got reduced.

Lasso Regression: The model built on Lasso regression at an optimal value of alpha has:

- Optimal value (alpha) = 0.001
- R2 value of train data is =0.92
- R2 Value of test data is = 0.91

Top features with coefficients:

Feature	Co-efficient
Total SF	0.75
TotalBsmtSF	0.28
OverallQual_9	0.22
GarageArea	0.18
OverallQual_10	0.17

In Lasso, the penalty is the sum of the absolute values of the coefficients. Lasso shrinks the coefficient estimates towards zero when alpha value increases thus helping in the variable selection.

After doubling the alpha value:

- (alpha) = 0.002
- R2 value of train data is =0.89
- R2 Value of test data is = 0.88

Top predictor variables after the alpha value is doubled:

Feature	Co-efficient
Total SF	0.74
TotalBsmtSF	0.27
GarageArea	0.19
OverallQual_9	0.18
BsmtFinSF1	0.12

The predictor variables are almost same but the coefficients got reduced.

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?

Ridge Regression: Ridge regression adds "squared magnitude" of coefficient as penalty term to the loss function. This regression regularization parameter is a scalar and it enforces the β coefficients to be lower but not zero i.e. it will not eliminate the features that are irrelevant but forces their coefficients close to zero making their impact on the model less. This technique works very well for overfitting issues.

Lasso Regression: Lasso Regression adds "absolute value of magnitude" of coefficient as penalty term to the loss function. This regression regularization parameter is an absolute value unlike ridge but this difference has a huge impact on the trade-off. Lasso is better by reducing the β coefficients and also makes them zero if the feature is irrelevant to the model.

The key difference being Lasso shrinks the less important features coefficients to zero thus, removing irrelevant features. So, this works well for feature selection in case we have large features.

In the model we built using Ridge and Lasso the overall accuracy for both Ridge and Lasso are almost same but since the model we built has large number of features including dummies we can prefer Lasso.

Out of 184 features Lasso picked only 70 features.

Out of 184 features only 6 features are almost equal to zero in ridge regression.

And the difference between the train accuracy and test accuracy is very less in Lasso than Ridge. So, Lasso is preferred over Ridge

	Ridge Regression	**Lasso Regression**
α	3	0.001
R2 value on Training Dataset	0.93	0.92
R2 value on Test Dataset	0.90	0.91

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?

The model built on Lasso regression at an optimal value of alpha has:

- Optimal value (alpha) = 0.001
- R2 value of train data is =0.92
- R2 Value of test data is = 0.91

Top features with coefficients:

Feature	Co-efficient
Total SF	0.75
TotalBsmtSF	0.28
OverallQual_9	0.22
GarageArea	0.18
OverallQual_10	0.17

After removing the top 5 features then:

The model built on Lasso regression at an optimal value of alpha has:

- Optimal value (alpha) = 0.0001
- R2 value of train data is =0.91
- R2 Value of test data is = 0.90

After removing top 5 features the next top features with coefficients:

Feature	Co-efficient
BsmtFinSF1	0.39
OverallCond_5	0.36
TotRmsAbvGrd_9	0.34
OverallCond_7	0.31
OverallCond_8	0.30

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:

If a model is stable and does not change drastically upon changing the training data then it can be considered as a robust model. By regularization we can make the model more stable. The model is considered generalisable if it does not overfits the training data, and works well with new data i.e. test data. Its implication in terms of accuracy is that a robust and generalisable model will perform equally well on both training and test data i.e. the accuracy does not change much for training and test data. AIC, BIC, R squared and advanced r squared values are some of the measures of the accuracy.