Problem Statement - Part II

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:

The optimal value of alpha in case of Ridge and Lasso is as below:
Ridge -100
Lasso - 0.01

If we double value of alpha, the coefficient values are decreasing. Also there are slight inclease in R2_score values.
Before:

Most important predictor variables are below:

	Ridge_coef	lasso_coef
GrLivArea	0.205396	0.300675
OverallQual	0.196527	0.234330
GarageCars	0.122654	0.114692
KitchenQual	0.100822	0.105293
ExterQual	0.094867	0.100958
BsmtExposure	0.077332	0.075494
BsmtQual	0.081840	0.068742
BsmtFullBath	0.047457	0.046377
MasVnrArea	0.056579	0.042261
MasVnrTypeenc	0.050409	0.038068
Foundationenc	0.049237	0.035663

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:

- The optimal alpha value in case of Ridge and Lasso is as below:
  Ridge -100
  Lasso - 0.01
- The metrics for Ridge and Lasso regression are:
  Ridge –
  r2_score train : 0.8274930979464847
  r2_score test : 0.8466496116030884
  RSS train: 168.24378497135933
  RSS test: 74.31776410692329
  MSE Train : 0.1647833349376683
  MSE Test : 0.16928875650779793

  Lasso -
  r2_score train : 0.8232158076530309
  r2_score test : 0.84806119421808
  RSS train: 172.41537173006586
  RSS test : 73.63367282489236
  MSE Train : 0.1688691202057452
  MSE Test : 0.16773046201570013

R2_score for train and test data I higher for Lasso.

Also, since Lasso helps in feature reduction as the coefficient value of some of the features became 0, Lasso has a better edge over Ridge.

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:

After removing top 5 predictor variables and building model again, below are the new 5 most important variables,

```
betas['lasso_coef'].sort_values(ascending=False).head(5)
```

```
2ndFlrSF        0.261615
TotalBsmtSF     0.225896
BsmtQual        0.134479
FullBath        0.116317
BsmtExposure    0.106243
Name: lasso_coef, dtype: float64
```

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:

Ensuring that a machine learning model is both robust and generalizable is crucial for its performance and reliability. Robustness and generalizability involve a trade-off. While robustness helps handle outliers and unexpected data, generalizability ensures reliable predictions on new data. Techniques like Regularization methods would help avoid overfitting then by making the model generalizable.

A robust model can exercise uncertain real-world scenarios appropriately and increase trust in the AI system. Achieving model robustness can imply slightly compromising accuracy to reduce generalization errors.

Generalization ensures the model's ability to make accurate predictions on new, unseen data.