Assignment based Subjective questions

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:
Optimal value for Ridge and lasso regression came out to be 20 and 0.0001 respectively according to the analysis.

After doubling:
Ridge- 40
Lasso- 0.0002

Top 3 predictors after doubling the alpha in ridge regression:

GrLivArea: 0.70725
2ndFlrSF : 0.059369
TotalBsmtSF: 0.049047

Top 3 predictors after doubling the alpha in lasso regression:

2ndFlrSF: 0.082957
GrLivArea : 0.065249
TotalBsmtSF: 0.057840

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 values of lambda for ridge and lasso regression came out to be : 20 and 0.0001.

```
Fitting 5 folds for each of 27 candidates
Optimum alpha for ridge is 20.000000
ridge  Regression with  20
===================================
R2 score (train) :   0.9204261317885009
R2 score (test) :   0.9071592670813748
RMSE (train) :   0.10351065012725395
RMSE (test) :   0.11158000823714954
```

```
]:    1  # setting the parameters for alpha to
      2  params = {'alpha': [0.000001, 0.00001
      3
      4  lasso_final_model, y_test_predicted

Fitting 5 folds for each of 12 candidates
Optimum alpha for lasso is 0.000100
lasso  Regression with  0.0001
===================================
R2 score (train) :   0.9210547430844427
R2 score (test) :   0.9050487594841023
RMSE (train) :   0.10310098687876074
RMSE (test) :   0.11284113103042774
```

Ridge is performing better in terms of R2 score in test dataset as well has lower RMSE in test dataset. Thus, we finalized ridge regression model.

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:

```python
# Printing the top 5 model coefficients in the new Lasso model
model_coeff = pd.DataFrame(index=X_test_new.columns)
model_coeff.rows = X_test_new.columns
model_coeff['Lasso'] = lasso_model.coef_
model_coeff.sort_values(by='Lasso', ascending=False).head(5)
```

	Lasso
BsmtFinSF1	0.096637
BsmtUnfSF	0.085716
LotArea	0.068313
ExterCond_Gd	0.052169
ExterCond_TA	0.051110

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:

The major challenge that we face with predictive models is that either the bias is high or the variance is high. To get the optimum value where there is a balance between bias and variance, we need to use regularization techniques such as Ridge or Lasso regression. By using these techniques, we can make sure that our model is robust, accurate and generalizable. In the case of overfitting, bias is very less and variance becomes very high. By using any regularization technique (Ridge or Lasso), we can enable a small increase in bias which in turn leads to a huge reduction in variance leading to robustness and generalizability of the model.