

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:

Technique	Optimal value
Ridge	0.01
Lasso	0.0001

When doubled the value for ridge:

```
1 # We will calculate some metrics such as R2 score, RSS and RMSE when Lambda=0.02
2 y_pred_train = ridge.predict(X_train)
3 y_pred_test = ridge.predict(X_test)
4
5 metric2 = []
6 r2_train_lr = r2_score(y_train, y_pred_train)
7 print("R-Squared (train set) with 2 X lambda for Ridge :",round(r2_train_lr*100,2),"%")
8
9 r2_test_lr = r2_score(y_test, y_pred_test)
10 print("R-Squared (test set) with 2 X lambda for Ridge :",round(r2_test_lr*100,2),"%")
11 print("\n")
```

R-Squared (train set) with 2 X lambda for Ridge : 94.51 %
R-Squared (test set) with 2 X lambda for Ridge : 83.85 %

When doubled the value for Lasso:

```
1 # We will calculate some metrics such as R2 score, RSS and RMSE when Lambda=10
2 y_pred_train = lasso.predict(X_train)
3 y_pred_test = lasso.predict(X_test)
4
5 r2_train_lr = r2_score(y_train, y_pred_train)
6 print("R-Squared (train set) with 2 X lambda for Lasso :",round(r2_train_lr*100,2),"%")
7
8 r2_test_lr = r2_score(y_test, y_pred_test)
9 print("R-Squared (test set) with 2 X lambda for Lasso :",round(r2_test_lr*100,2),"%")
10 print("\n")
```

R-Squared (train set) with 2 X lambda for Lasso : 88.98 %
R-Squared (test set) with 2 X lambda for Lasso : 85.97 %

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: Lasso is chosen as it also helps in the feature elimination and the r^2 score on train and test is within a range of 5% which shows the model is decent in prediction.

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: These are the following variables after performing the operation:

	Feature	Coeff
11	TotalBsmtSF	0.298081
13	2ndFlrSF	0.209232
3	LotArea	0.085126
7	MasVnrArea	0.061773
15	FullBath	0.051928

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: Model is robust as the R^2 score for train and test is close and the over fitting has been removed along with the unwanted features using Lasso regression.