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 ?

The optimal values are :

Optimal values for Ridge {'alpha': 100}
Optimal values for Lasso {'alpha': 0.01}

The metrics for data comparison

Metric	Ridge Regression	Lasso Regression
R2 Score (Train)	0.925907	0.916897
R2 Score (Test)	0.838708	0.840066

When we double the value of alpha for Ridge (200) and Lasso (0.02) there will change in R2 score of train data as shown below.Also the coefficient value changes. But overall performance remains the same.

Metric	Ridge Regression	Lasso Regression
R2 Score (Train)	0.922085	0.905198
R2 Score (Test)	0.837291	0.837553

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?

The optimal values are :

Optimal values for Ridge {'alpha': 100}
Optimal values for Lasso {'alpha': 0.01

Since the performance of Ridge and Lasso is almost the same with respect to R2 ~83% and Mean square error ~0.47. We will choose Lasso since it will help in feature elimination by giving a penalty and making coefficients close to 0.

Metric	Ridge Regression	Lasso Regression
MSE (Train)	0.272201	0.288276
MSE (Test)	0.478352	0.476334

Question 3
After building the model, you realized 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 top 5 predictor variables for Lasso are :
'GrLivArea','OverallQual','LotArea','BsmtFinSF1','GarageArea'

If we exclude them and run the model then,

The top 5 predictor variables becomes :
2ndFlrSF,1stFlrSF,TotalBsmtSF,Neighborhood_NridgHt,BsmtExposure

Also the R2 score changes for test data from 0.84 to 0.81

r2_train : 0.8576749168226148
r2_test : 0.8172575158024807
rss1 : 143.46368384280424
rss2 : 112.2561192990664
mse_train : 0.14232508317738515
mse_test : 0.259252007619091

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?

To make model robust and generalisable we follow the Occam's Razor principles which suggests :
Given 2 models that show similar performance in the finite training/test data we should pick the one that makes fewer assumptions for the unseen data.

- Simple model are more generic and widely applicable
- Simple model require less training sample then complex one so are easier to train
- Simple models are more robust:
    - Simple model has low variance high bias
    - Complex model lead to overfitting

Regularization is the process in machine learning to simplify models. Regularization helps to strike balance between keeping the model simple yet not making it too simple to be of any use.

This raises the question how much error is willing to tolerate during training to gain generalization. The simplification can be done on : choice of simpler functions,keep the number of model parameters small,keeping the degree of polynomials low.Regularization is the simplification done by training algorithms to control.

The accuracy of the model can be maintained by keeping the balance between Bias and Variance .
Bias is how accurate the model is likely to be on future (test) data.
Variance is the variance in the output on test data with when there is change in training dataset.

The low complexity models have high bias and low variance.and high complexity models will have low bias high variance. The best model is one that balances both achieving a reasonable degree of predictability i.e low variance without too much compromise on the accuracy i.e bias.

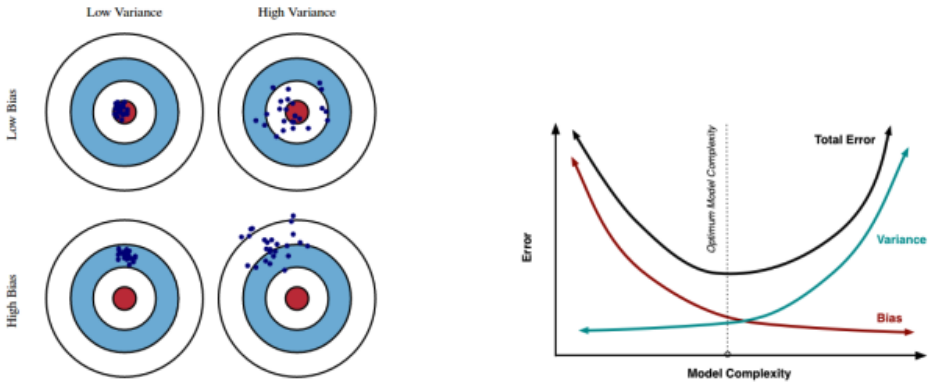

Figure 2: (Left) Illustration of Bias and Variance; (Right) Bias-Variance Tradeoff