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 of alpha for

- ridge is 1.0
- lasso regression is 0.001

When we double the alpha i.e.,
ridge is 2.0 then R2 values of train reduces slightly and increases for test but RSS values have similar behaviour when alpha is 2.0

For lasso 0.002, R2 value for train and test reduces by 1 when compared to 0.001

```
Lasso
Lasso(alpha=0.002)
```

```
# Lets calculate some metrics such as R2 score, RSS and RMSE

y_pred_train_lasso = lasso.predict(X_train)
y_pred_test_lasso = lasso.predict(X_test)

metric3 = []
r2_train_lr = r2_score(y_train, y_pred_train_lasso)
print(r2_train_lr)
metric3.append(r2_train_lr)

r2_test_lr = r2_score(y_test, y_pred_test_lasso)
print(r2_test_lr)
metric3.append(r2_test_lr)

rss1_lr = np.sum(np.square(y_train - y_pred_train_lasso))
print(rss1_lr)
metric3.append(rss1_lr)

rss2_lr = np.sum(np.square(y_test - y_pred_test_lasso))
print(rss2_lr)
metric3.append(rss2_lr)

mse_train_lr = mean_squared_error(y_train, y_pred_train_lasso)
print(mse_train_lr)
metric3.append(mse_train_lr**0.5)

mse_test_lr = mean_squared_error(y_test, y_pred_test_lasso)
print(mse_test_lr)
metric3.append(mse_test_lr**0.5)
```

```
0.9226888559450409
0.8926125956395121
10.371740919629087
6.49594043771276
0.011706253859626509
0.017094580099244108
```

R2 is better for Ridge regression after implemenation hence the predictor variables considering coefficients are:
- LotArea
- OverallQual
- OverallCond
- YearBuilt
- BsmtFinSF1
- TotalBsmtSF
- 1stFlrSF
- 2ndFlrSF
- GrLivArea
- GarageArea
- ScreenPorch

We can see some new predictor varibales compared to Lasso regression with 0.001 alpha(our best model).
But considering Ridge doesn't do feature tuning, the predictor variables would be same as Lasso with different coefficients after feature tuning and multicolinaerity handling.

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: R2 score of Lasso Regression for test data is slightly higher(to a hundredth decimal) than Ridge regression so I cloose Lasso Regression.

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: The five most important predictor variables now are:

1. OverallCond
2. 1stFlrSF
3. 2ndFlrSF
4. TotRmsAbvGrd
5. GarageArea

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 model should be built by proper feature tuning and avoid multipcolinearity but no so much that we underfit. Thus making it robust and generic..
- The implications of the same can be seen when we see accurate prediction for datasets other than just the training dataset thus improving accuracy.This is important because we always built for unseen real world data and expect high accuray for the same,