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?

Optimal Value for Ridge is 2

Optimal Value for Lasso is .01

When we increase the value of the hyperparameter alpha in Ridge and Lasso. Below is the effect:

In Lasso More features will be pushed to 0 as the penalty applied is high.

In Ridge more coefficients will be pushed towards zero.

On Doubling the value for Lasso:

```
    sorted(list(zip(cols, model_parameters)),key=(lambda x:-(x[1])))

[769]  ✓ 0.4s

...  Output exceeds the size limit. Open the full output data in a text editor
     [('constant', 12.031),
      ('OverallQual', 0.13),
      ('GrLivArea', 0.102),
      ('GarageArea', 0.049),
      ('Fireplaces', 0.025),
      ('OverallCond', 0.021),
      ('CentralAir_Y', 0.02),
      ('BsmtFullBath', 0.019),
      ('TotalBsmtSF', 0.018),
      ('Foundation_PConc', 0.011),
      ('LotArea', 0.008),
      ('WoodDeckSF', 0.007),
      ('BsmtFinType1_GLQ', 0.006),
      ('Neighborhood_Crawfor', 0.005),
      ('Neighborhood_NridgHt', 0.005),
      ('FullBath', 0.004),
      ('Condition1_Norm', 0.003),
      ('BsmtExposure_Gd', 0.003),
      ('LotConfig_CulDSac', 0.002),
      ('MasVnrArea', 0.0),
      ('BsmtFinSF1', 0.0),
      ('BsmtFinSF2', 0.0),
      ('BsmtUnfSF', -0.0),
      ('2ndFlrSF', 0.0),
      ('LowQualFinSF', -0.0),
```

On Doubling the value for Ridge:

```
    cols = cols.insert(0, "constant")
    sorted(list(zip(cols, model_parameters)),key=(lambda x:-(x[1])))

[777]  ✓  0.1s

...  Output exceeds the size limit. Open the full output data in a text editor
     [('constant', 12.031),
      ('RoofMatl_CompShg', 0.259),
      ('RoofMatl_Tar&Grv', 0.179),
      ('RoofMatl_WdShngl', 0.124),
      ('GrLivArea', 0.122),
      ('RoofMatl_WdShake', 0.103),
      ('MSZoning_RL', 0.087),
      ('MSZoning_RM', 0.068),
      ('RoofMatl_Metal', 0.063),
      ('RoofMatl_Roll', 0.062),
```

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?

We will make use of Lasso Regression as it will help in feature elimination as well. By reducing the coefficients of the features to 0

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?

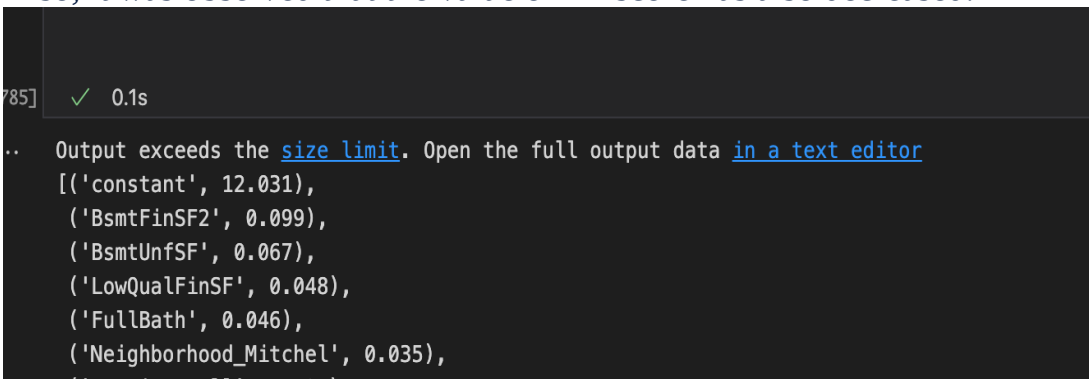
The 5 most important feature are as below:

1. OverallQual
2. GrLivArea
3. GarageArea
4. OverallCond
5. BsmtFullBath

If we drop these features then, the top 5 features are:

1. BsmtFinSF2
2. BsmtUnfSF
3. LowQualFinSF
4. FullBath
5. Neighborhood

Also, it was observed that the value of R2 score has also decreased.



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?

Going by Occam's Principle the model should be as simple as possible. Variance should not have affect on the model and it should be able to predict the correct result. For this we have to take care that variance is not high in the model. If the variance is high then we can say that model has overfitted the data.

The only consequence of a simplified model is that it will have less accuracy. As a complex model will have high accuracy and high Variance. Since now variance is less for the model the bias will increase to a certain extent thereby decreasing the accuracy of the model