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 of alpha for ridge: 1.0

Optimal value of alpha for lasso: 0.0001

If we choose to double the value of alpha for both ridge and lasso:

In case of ridge that will lower the coefficients and in case of Lasso there would be more less important features coefficients turning 0

The most important predictor variables after the change is implemented are those which are significant,

1. LotFrontage
2. BsmtFullBath
3. OverallCond
4. CentralAir
5. OverallQual

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 got good score for both the models so we can go with Lasso Regression as it results in model parameters such that lesser important features coefficients become zero.

Ridge Train:0.93 Ridge Test: 0.75

Lasso Train:0.93 Lasso Test:0.75

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 important predictor variables now?

1. LotArea
2. FullBath
3. ExterCond
4. 1stFlrSF
5. MSZoning_RH

Question 4:

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

1. Use a model that's resistant to outliers
2. Use a more robust error metric