Q1: 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?

A1: The optimal value of alpha for ridge and lasso regression is as follows:

Ridge: 7.0

Lasso: 0.001

The R2 value obtained for both of them are as follows:

Ridge:

Train: 0.937

Test: 0.924

Lasso:

Train: 0.92

Test: 0.925

The top 5 predictor variables are:

Ridge:

1. 2ndFlrSF
2. OverallCond_8 (Very Good)
3. SaleCondition_AdjLand
4. OverallQual_8 (Very Good)
5. OverallQual_7 (Good)

Lasso:

1. OverallQual_8 (Very Good)
2. 2ndFlrSF
3. OverallQual_7 (Good)
4. Neighborhood_CollgCr (College Creek)
5. Functional_Sev (Severely Damaged)

Following are the changes if we double the values of Ridge and Lasso, which is 14 and 0.002.

1. R2 Score
    a. Ridge:
        i. Train: 0.933
        ii. Test: 0.926
    b. Lasso:
        i. Train: 0.904

ii. 0.911

The values of R2 score for ridge is almost the same but the gap between test and train r2 increases slightly and hence bias slightly increases. For Lasso the values of Train and test R2 score reduces and hence the original values of 7 for ridge and 0.001 for lasso is correct.

2. The predictor variable also change:
    a. Ridge (Sale Condition AdjLand drops while Neighbourhood_CollgCr come in)
        i. 2ndFlrSF
        ii. OverallQual_7
        iii. OverallQual_8
        iv. Neighbourhood_CollgCr
        v. OverallCond_8
    b. Lasso (the order of importance of coefficients applied changes, the list remains the same)
        i. OverallQual_8
        ii. 2ndFlrSf
        iii. OverallQual_7
        iv. Neighbourhood_CollgCr
        v. Functional_Sev

Q2: 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?

A2: From the R2 scores the test score for both of them is almost the same. Also, since there is almost no difference between the r2 score between ridge and lasso regression both the regression model show same amount of bias. The value of lasso is slightly higher, also the train and test scores for lasso are almost the same which shows low variance as well. But, since lasso reduces a lot of features to 0 and hence the number of total features are much lower (81 compared to 257 for ridge) the model is much simpler and between a model whose r2 score is almost same we will always try and choose the simpler model which will reduce the variance.

Q3: 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?

A3: The five most important predictor variables in lasso regression are:
'OverallQual_8','2ndFlrSF','OverallQual_7','Neighborhood_CollgCr','Functional_Sev'

After removing those the alpha value reduces to 0.0001 and the R2 score also reduces to 90.9 for test dataset.

The five mot important predictor variables now are:

1. SaleType_ConLI
2. OverallCond_4
3. 1stFlrSF
4. Neighborhood_Sawyer
5. Neighborhood_BrkSide

Q4: 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?

A4: A robust model means that any variation in the data does not affect the performance of the model. The variation between training data and any new test data would be low for a robust model.

A generalizable model means that the model is able to adapt to new, previously unseen data or in other words it show low variance.

To make sure that a model is robust and generalizable, we need to make sure that the model does not overfit on the test data. Overfitting induces a very high variance in the model even though the bias is very low. Any small change in the data will wildly change the model's performance. One way of making sure that the model does not overfit is by making it less complex by introducing regularization.

A highly complex model will have very high accuracy. So, to make our model more robust and generalizable, we will have to decrease variance which will lead to some bias. Addition of bias means that accuracy will decrease.