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?

Ans:- Optimal value of alpha is value which minimizes the sum of bais and variance is minimum.

In housing data set the optimal value of alpha for lasso:- 0.001
In housing data set the optimal value of alpha for ridge:- 100

If double value of alpha for ridge and lasso: in case of ridge it will lower the the coefficient and in case of lasso for insignificant predictor coefficient will tend to zero.

Most important predictor variables after the changes will be the variable which are more significant .

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?

Ans:

optimal value for lambda for Lasso : {'alpha': 0.001}
Lasso:- Train R2 Square : 0.94 Test R2 Square : 0.87

optimal value for lambda for Ridge {'alpha': 100}
Ridge:- Train R2 Square : 0.93 Test R2 Square : 0.88

As we got good score in both Lasso and ridge will go with Lasso because in lasso insignificant predictor will be zero.

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?

Ans:
5 most significatnt feature in Lasso

PoolArea 0.220
PoolQC_Fa 0.199
GrLivArea 0.111
PavedDrive_Y -0.070
OverallQual 0.066

5 most significatnt feature in Ridge

OverallQual 0.054
GrLivArea 0.044
PavedDrive_Y -0.036
OverallCond 0.033
1stFlrSF 0.031

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?
Ans:

The model should be as simple. There should be Bais-variance tradeoff . the simple model will be more bias but less variance and more general.
Bais is error in model when model is weak to learn from the data. high bias means model is weak.
variance : this also is error in model . when model is overfitted. high variance means its not generalized.
so there should be trade of between bias and variance.
Also we need to handle outlier . it should be removed from data set.