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

The optimal value of alpha for Ridge is 1 and for Lasso it is 0.001.

If we chose to double the value of alpha :
In Ridge, mean squared error slightly increase but the r2 values remains almost same for test and train.
In Lasso, mean squared error increases slightly but the r2 values decreases for both train and test.

The most important predictor variables are

For Ridge:

1. Total_sqr_footage
2. OverallQual
3. GrLivArea
4. Neighborhood_StoneBr
5. OverallCond
6. TotalBsmtSF
7. LotArea
8. YearBuilt
9. Neighborhood_Crawfor
10. Fireplaces

For Lasso:

1. Total_sqr_footage
2. OverallQual
3. YearBuilt
4. GrLivArea
5. Neighborhood_StoneBr
6. OverallCond
7. LotArea
8. Neighborhood_Crawfor
9. Neighborhood_NridgHt
10. GarageCars

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 :
The optimal value of LAMBDA we got in case of Ridge and Lasso is :

Ridge - 1.0
Lasso - 0.0001
The r2 value we got in case:

Ridge - Train = 0.938 , Test = 0.846
Lasso - Train = 0.935 , Test = 0.828
The Mean Squared error in case of Ridge and Lasso is:

Ridge - 0.0051
Lasso - 0.0057
We can clearly observe that the Mean Squared Error of Lasso is slightly higher than that of Ridge. But the difference is minimal.

Lasso helps helps in feature reduction and helps in model interpretation by taking the coefficients .So according to me Lasso is a better option.

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?
Ans:

1.LotArea
2.TotalBsmtSF
3.OverallCond
4.TotRmsAbvGrd
5.Total_Bathrooms

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 simple or generic. The accuracy might decrease but it will be more robust.
According to Bias-Variance trade-off, simpler the model more the bias but less the variance. A complex model will have low bias but more variance.
There is always a list of overfitting and underfitting in a model.
Using Regularization, we can reduce the chance of overfitting as it shrinks the coefficients to zero and makes the model less complex.
It penalises the model if it becomes more complex.