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?

Answer:

1. The Optimal Value for Alpha in Lasso Regression is 0.001
2. The Optimal Value for Alpha in Lasso Regression is 5
3. Higher the value of Alpha, the lower the number of nonzero coefficients. The number of features would be reduced from 33 features.
4. Two most important predictors are: GrLivArea and OverallQual with highest absolute value of beta

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?

I would prefer Lasso Regression as I can shortlist the best features using that in an automated fashion and I would know the best 33 features to go with.

Question 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?

In that case, the variables with highest absolute value of coefficients would be the 5 most important features.

It would be - Fireplaces, Full Bath, Neighbourhood (Northridge Heights), Neighborhood (NorthRidge), Basement Exposure.

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?

We can make sure by not overfitting the model. The R2 Score should not be too high in Training Data, making sure that we have a flexible and robust model.

Higher the overfitting, lower the accuracy - so if you go for a robust, flexible and generalisable mode, you will be able to get a more accurate model.