Subjective Questions

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 for Ridge Regression is '50' and for Lasso Regression is '20'
If the alpha is more than the optimal value then the model gives more importance to the variable and it may result in overfitting.
The top 5 predictor variables after the value of alpha is doubled are:

Ridge Regression with alpha = 100:

OverallQual
BsmtQual_Ex
Neighborhood_NoRidge
Neighborhood_NridgHt
BsmtExposure_Gd

Lasso regression with alpha = 40:

Condition2_PosN
RoofMatl_WdShngl
RoofMatl_CompShg
Neighborhood_NoRidge
RoofMatl_WdShake

There were some changes in summary statistics but, the MSE value increased considerably when the alpha value was doubled.

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: We judge any regression model by its summary statistics and among those 'r2' has even more weightage.
Below are the summary stats for different model:

Metric	Linear Regression	Ridge Regression	Lasso Regression
R2 Score(Train)	9.484347e-01	8.717502e-01	9.374591e-01
R2 Score(test)	-2.931258e+08	8.637399e-01	8.450176e-01
RSS(Train)	3.290227e+11	8.183234e+11	3.990547e+11
RSS(test)	8.262374e+20	3.840781e+11	4.368510e+11
MSE(Train)	1.795147e+04	2.831064e+04	1.976985e+04
MSE(test)	1.373458e+09	2.961234e+04	3.158127e+04

Here the 'r2' for train and test is very close for Ridge regression, also by seeing the r2 for Lasso regression, it seems to have overfit slightly compared to Ridge.
MSE stats for test and train also favor Ridge. So I would choose the Ridge regression model.

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:
Top 5 predictors before we discard are:
RoofMatl_WdShngl
Condition2_PosN
RoofMatl_CompShg
RoofMatl_Membran
RoofMatl_WdShake

Top5 predictors after the above predictors:

Neighborhood_NoRidge
BsmtQual_Ex
Neighborhood_NridgHt
Neighborhood_Somerst
Neighborhood_Crawfor

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:

Model is a mathematical tool developed to predict the target based on past data, relation between predictors and target. So we need the model to be robust as well as sensitive. This can be achieved by developing an intelligent cost function.
Since we are trying to predict, we are dealing with approximates rather than absolutes.

There are two major issues in building a model:

1) Overfitting: When a model performs well on training data but fails to explain on test data.
2) Underfitting : When the model fails to explain the train data itself.

We need a well balanced model which is not too complex(overfit) or too simple(underfit).To balance this there are two ways to improve the MLR, two of them are Ridge and Lasso regression.
In these Regression methods, the accuracy improving K-fold techniques and hyperparameters selection methods are incorporated which helps in regularizing the effect of any single predictor so as to not overfit but maintain a better balance.