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:

Currently value of alpha for both Lasso and Ridge regression is .001 and .05 respectively. Once we double the value of alpha for Lasso (.002) and Ridge (10.0) found following changes.

Lasso Regression Analysis:

Once value of alpha doubled, I saw there was change in order of predictive variable:
When Alpha: 0.002, top 5 predictive variable.
ScreenPorch
1stFlrSF
2ndFlrSF
MSZoning_FV
BsmtFullBath

When Alpha: 0.001, top 5 predictive variable.
ScreenPorch
MSZoning_FV
BsmtFullBath
2ndFlrSF
1stFlrSF

Ridge Regression analysis:

Once value of alpha doubled, I saw there is change top 5 predictive variable:

When alpha : 10.0 , top 5 predictive variables.
1stFlrSF
2ndFlrSF
BsmtFullBath
MiscVal
MSZoning_FV

When alpha : 5.0 , top 5 predictive variables.
ScreenPorch
GarageCond
EnclosedPorch
3SsnPorch
MSZoning_FV

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:

I will choose Lasso Regression model because it will penalize the data set in case of feature redundancy and also takes care of feature elimination. As a result of this we can avoid overfitting a model on dataset.

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: Next five available variable available in order will become predictive variable for model.

Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Ans: In order to make the model accurate we need to first filter out the outlier so that it outliers will get weight in analysis and train and test accuracy will be maintained thereby accuracy predicted by model will be high. And we should be using Keep checking the number of features to be included in model, basically multicollinearity check should be there, for this purpose Lasso regression is a suitable choice which takes care of redundant feature elimination and adds penalty on feature addition.

If model is not generalizable it may affect the accuracy of model and results will drastically change once tested of unseen data or test data.