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:

Optimal Value of alpha for ridge regression is: 500, optimal value for lasso regression is 500.

If we double the value of alpha for both ridge and lasso regression, then R2-squared for train set of lasso regression becomes 87.33 from 90.47, similarly the R2-squared for train set of ridge regression becomes 84.00 from 86.63. The most important predictor variables after the changes are: OverallCond, BsmtFullBath, NridgHt.

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?

Answer 2:

I will choose lasso regression because its R2-squared is good as compared to R2-squared of ridge regression, which is a good indicator for model fit. Also, Lasso regression help in elimination of features.

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?

Answer 3:

(a) StoneBr
(b) GarageArea
(c) FirePlaces
(d) 2ndFlrSf
(e) HalfBath

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?

Answer 4:

If we want to increase the accuracy of our model:

(a) We have to treat missing and outliers values
(b) Add more data into the dataset because presence of more data results in better and accurate models.
(c) We have to select features based on various metrics.
(d) Cross validation is the most important factor of increasing accuracy of the model.

Our model is robust and generalisable if it is not affected by ordinary least squares.