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:

Optimal value of Alpha	Ridge	Alpha = 0.1	R2 = 0.88
	Lasso	Alpha = 0.001	R2 = 0.85
Double the value of alpha	Ridge	Alpha = 0.2	R2 = 0.78
	Lasso	Alpha = 0.002	R2 = 0.87

After doubling the alphas, the predictor variables as follows.

Ridge Alpha Co-Efficient after Do	
MSZoning_RL	0.3
MSZoning_FV	0.3
MSZoning_RH	0.3
MSZoning_RM	0.3
RoofMatl_Membran	0.
RoofMatl_WdShngl	0.2
RoofMatl_Roll	0.2
Condition2_PosA	0.2

Lasso Co-Efficient after Doub	
OverallQual	0.102
GrLivArea	0.101
Neighborhood_Crawfor	0.068
Neighborhood_NridgHt	0.057
Neighborhood_Somerst	0.055
Condition1_Norm	0.050
GarageCars	0.050
YearBuilt	0.049

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: Based on R2 score, Ridge model is better. Mean square error is almost same for both models.

Will be choosing Lasso model because if offers feature reduction.

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: Skipped.

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?
   - The benefits of generalizing the model is the Test accuracy does not change erratically compared with training set. Upon, checking with unknown test sets, the accuracy will be usually approximate to that of training set.
   - A model can be made more robust and generalisable by using additional experiments, such application of cross validation, properly choosing hyper parameter along with application of proper penalty ensuring mean squared error to be minimum.
   - Outliers may affect the accuracy of the given model. The outliers should be analyzed , and based to analysis it can be retained if they are relevant to data . Or dropped if they are not relevant.