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 most optimal value for ridge regression is 5 and that for lasso regression is 0.0005.
When we increase alpha in ridge and lasso regression, it means that we are giving a little more priority to the regularization term when compared to the error term in the cost function. When we increase the term, we allow the model to make a little more error which making it more generalizable. In other words, we increase the bias and decrease the variance of the model.
Actually increasing alpha by a factor of two would cause a slight decline in the train scores as the model attempts to generalize better. The original train scores for ridge and lasso are 0.9190 and 0.9186. The new train scores are 0.9181and 0.9167 respectively.

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:
Lasso Regularization - On applying the optimal value of lambda, the coefficients of redundant variables becomes zero . Hence Lasso indirectly performs feature selection on the dataset.
Ridge Regularization - On applying the optimal value of lambda , the coefficients of redundant variables becomes close to zero but does not exactly zero. Thus Ridge does not perform feature selection on the dataset.

For this specific case, both the ridge and lasso give similar results though. We could leverage ElasticNet which gives us the capabilities of both ridge and lasso regression. However, lasso regression would be a better choice as the range of coefficients that we get in lasso is wider than ridge and also it takes care of correlated variables (which are plenty) and reduces them to zero

As the number of variables in the dataset provided is very high, Lasso Regression is more preferable as it performs feature selection.

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:

The 5 most important factors in the current lasso model were:
'MSZoning_RL'
'GrLivArea'
'OverallQual'
'MSZoning_RM'
'MSZoning_FV

After removal of these parameters and rebuilding the model, we get the following top 5 positively and top 5 negatively correlated variables.

	Coefficient	Column
0	-0.055114	Exterior1st_VinylSd
1	-0.028950	Exterior1st_Wd Sdng
2	-0.021431	Neighborhood_OldTown
3	-0.021288	Exterior2nd_Stucco
4	-0.020384	IsRemodelled

	Coefficient	Column
40	0.046540	Foundation_PConc
41	0.053953	OverallCond
42	0.055784	TotalBsmtSF
43	0.085613	1stFlrSF
44	0.100625	2ndFlrSF

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:
The model can be made robust and generalizable by applying Regularization on the model.
Accuracy for Train Data:-
With Increase in value of hyper parameter lambda of Regularization, the error term increase constantly resulting in continually decrease in accuracy
Accuracy for Test Data:-
With Increase in value of hyper parameter lambda of Regularization, the error term decrease initially and then increase constantly. Hence Accuracy will increase initially and then decrease constantly (as the hyper parameter increases).
Accuracy for Overall Model:-
With Increase in value of hyper parameter lambda of Regularization, the error term increase constantly. Hence Accuracy will decrease constantly (as the hyper parameter increases).