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:

Optimal value of alpha for Lasso regression: 0.001

Optimal value of alpha for Ridge regression: 25

Most important predictors before change:

Lasso	**Ridge**
GrLivArea	OverallQual
OverallQual	GrLivArea
GarageCars	GarageCars
OverallCond	OverallCond
YearBuilt	YearBuilt

After the alphas are doubled: (the predictors remained the same)

Lasso	**Ridge**
GrLivArea	OverallQual
OverallQual	GrLivArea
GarageCars	GarageCars
OverallCond	OverallCond
YearBuilt	YearBuilt

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:

I will prefer the Lasso regression model because it will determine which parameters are to be used and the lasso regression model makes the coefficients of unimportant predictors to zero. In that way, it is very clear that we need not include those variables in further consideration

---

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:

We will ignore the first 5 variables and redo the same analysis and pick the next 5 variables.

Doing the same, these are the next 5 predictor variables:

Lasso	**Ridge**
PoolQC	2ndFlrSF
2ndFlrSF	1stFlrSF
1stFlrSF	KitchenQual
ExterQual	ExterQual
KitchenQual	GarageArea

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:

We must ensure that the model is not overfit or underfit. We have to ensure that the training score and test score are high enough, but the difference between them should be less.

A model that performs reasonably well with training and test data is expected to run a better analysis. That is the reason we use cross validation to verify the generated model.

Such model will have a higher bias which leads to a compromise in accuracy, but is more resilient to newer data because it is not underfit or overfit.