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?

Optimal values are

Ridge – 0.3
Lasso – 0.0001

Not much difference in terms of r2 score after doubling the alpha value. The predictor variables remain same but with different coefficients.

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?

The performance of both the models were nearly same in terms of MSE. But lasso helps in feature selection or reduction, so it is better to choose lasso.

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?

Following were the top 5 features:

	Feaure	Coef
0	MSSubClass	11.998372
4	OverallCond	0.139163
9	BsmtUnfSF	0.133010
22	GarageArea	0.121907
26	YrSold	0.120739

After removing and rebuilding the model, the top 5 features are :

	Feautre	Coef
0	LotFrontage	11.804384

	Feautre	Coef
10	FullBath	0.368331
9	BsmtFullBath	0.348788
8	GrLivArea	0.302298
11	HalfBath	0.287586
20	MoSold	0.203106

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?

To make it more generalisable and accurate, we should have a balance of bias and variance.

As a result of this, the accuracy on training set decreases but performance on test set
increases.