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?

There is no any optimal value of alpha for ridge and losso regression it depends upon the model Tuning the parameter alpha helps us to determine the how much we wish to regularize the model.

If we choose the double value of alpha for ridge module coefficients reduces to half values that means higher the value of alpha lower the value of model coefficients.

Living area in model is more important predictor variable in the model.

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?

We want to use losso regression for following reasons

Ridge regression retains all the variables that are present in the data. when the number of variables is very large and the data may have unrelated or noisy variables, we may not want to keep such variables in the model. Lasso regression helps us here by performing feature selection. And lasso regression shrinks the coefficients estimates towards zeo.

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?

Fireplace,

Garageyearbuilt

garageCars

woodenDeckSF

TotalRMSabovegrd.

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?

Lasso should perform better in situations where only a few among all the predictors that are used to build our model have a significant influence on the response variable. So, feature selection, which removes the unrelated variables, should help. But Ridge should do better when all the variables have almost the same influence on the response variable.