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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Ans: $Lasso_{\alpha} = 0.01$ and $Ridge_{\alpha} = 10$

Making alpha double for **Lasso** make the model simpler and total number of features with non-zero coefficients become 11 form earlier 18 with little decrease in accuracy / R^2 from 0.86 to 0.84. For **Ridge** also accuracy / R^2 decreases to 83.7 from 85 for same number of features (11) as Lasso but Ridge accuracy increases by 1% if we keep the features (18) as earlier while double the alpha but making it simpler by reducing features to 11 as Lasso with help of RFE decreases accuracy.

Important Variables for Lasso now: Important Variables of Ridge now:

R^2 =0.8430

GrLivArea	0.1655
TotalBsmtSF	0.0665
BsmtFinSF1	0.0340
GarageArea	0.0267
LotArea	0.0119
OpenPorchSF	0.0016
MasVnrArea	0.0014
ScreenPorch	0.0002
FireplaceQu_NA	-0.0112
YearSinceRemodAdd	-0.0589
HouseAge	-0.0837

R^2 =0.8377

GrLivArea	0.200
OverallQual_9	0.138
Neighborhood_Crawfor	0.130
BsmtExposure_Gd	0.115
CentralAir_Y	0.108
GarageCars_3	0.103
Functional_Typ	0.092
ExterQual_TA	-0.093
HouseAge	-0.122
KitchenAbvGr_2	-0.127
MSSubClass_160	-0.161

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?

Ans: Lasso would be chosen over ridge due to its accuracy / R^2 = 0.86 more than Ridge (R^2 = 0.85) for same number of features = 18. Also, Lasso has automatic feature selection capability and does not require RFE for 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?

Ans: Below are the top 5 variables when model created again excluding 5 important features.

Features	Weight
1stFlrSF	0.190
2ndFlrSF	0.124
BsmtFinType1_GLQ	0.021
GarageCars_3	0.020
MasVnrArea	0.017

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?

Ans : To make a model generalize better we need to keep it simple enough to avoid overfitting and but not so simple that it is not able to learn enough from the training data. It is a point from where test score start decreasing and gap between train and test score also start increasing or become constant.

0.01 is right value of alpha to keep accuracy above 80% and at the same time keeping model simple to generalize better. For alpha = 0.001 test accuracy would be more but mode will not be simple enough as will select more number of features to predict that level of accuracy(90%).