Question1: 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: Currently optimal value of alpha for ridge is 10 and lasso is 0.0001

If I doubled the alpha for ridge value will be

```
Model Evaluation : Ridge Regression, alpha=20.0
R2 score (train) :  0.9163
R2 score (test) :  0.8702
RMSE (train) :  0.1133
RMSE (test) :  0.154
```

If I doubled lasso values will be

```
Model Evaluation : Lasso Regression, alpha=0.0002
R2 score (train) :  0.9165
R2 score (test) :  0.8702
RMSE (train) :  0.1132
RMSE (test) :  0.154
```

Most predicted 50 variables are:

MSSubClass,LotArea,LandSlope,OverallQual,OverallCond,YearBuilt,BsmtQual,BsmtExposure,BsmtFinSF1,HeatingQC,CentralAir,1stFlrSF,2ndFlrSF,BsmtFullBath,KitchenQual,Functional,Fireplaces,GarageFinish,GarageArea,GarageQual,OpenPorchSF,MSZoning_RL,Street_Pave,Neighborhood_Edwards,Neighborhood_Names,Neighborhood_NWAmes,Neighborhood_NridgHt,Neighborhood_Somerst,Condition1_Feedr,Condition1_Norm,Condition2_Norm,BldgType_TwnhsE,RoofStyle_Gable,RoofStyle_Hip,Exterior1st_HdBoard,Exterior1st_Plywood,Exterior1st_Wd Sdng,Exterior2nd_HdBoard,Exterior2nd_Wd Sdng,MasVnrType_BrkFace,MasVnrType_None,MasVnrType_Stone,Foundation_PConc,Heating_GasA,GarageType_Attchd,GarageType_Detchd,GarageType_Not_applicable,PavedDrive_Y,SaleCondition_Normal,SaleCondition_Partial

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?

Comparing the values of R2 both the values are approx. same on train lasso is little better so I will choose that.

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:

	Ridge (alpha=10.0)	Lasso (alpha=0.0001)	Ridge (alpha = 20.0)	Lasso (alpha = 0.0002)
1stFlrSF	0.125766	0.129201	0.122390	0.128886
2ndFlrSF	0.113245	0.116214	0.110027	0.115560
OverallQual	0.077966	0.077953	0.078152	0.078312
OverallCond	0.050080	0.051006	0.049158	0.050966
YearBuilt	0.044336	0.046701	0.042158	0.046528

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

The model should be as simple as possible because this will increase its robustness and generalization but it will reduce accuracy. The Bias-Variance trade-off can also be used to understand it. The bias increases with model complexity while decreasing variance and it will increase generalization. Its accuracy implication is that a robust and generalizable model will perform similarly on both training and test data, i.e., the accuracy does not change significantly for training and test data.

Bias: When a model is unable to learn from the data, it makes mistake. High bias prevents the model from learning specifics from the data.

Variance: Variance is a model error that results from the model trying to learn too much. High variance refer to the model performs remarkably well on training data since it was well trained on those data, but it performs dreadfully on testing data because that data was unknown to the model.

To prevent overfitting and underfitting of data, bias and variance must be balanced.