Subjective Questions

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?

The optimal values of alpha for ridge and lasso regression are 1.7 and 0.001 respectively. After doubling the value of alpha for both ridge and lasso our R2 Score remains same for both train and test which is approx 0.88. We have the same R2_score for both lasso and ridge. Top features for both are listed below:

Best Lasso

	Feature	Coef
46	PoolArea	0.427682
77	Neighborhood_SWISU	0.388364
32	KitchenQual	0.381370
12	BsmtCond	0.349738
31	KitchenAbvGr	0.326874
30	BedroomAbvGr	0.265563
2	LotShape	0.214158
13	BsmtExposure	0.164857
19	TotalBsmtSF	0.156727
5	OverallCond	0.148813

Doubled Lasso

	Feature	Coef
46	PoolArea	0.654362
77	Neighborhood_SWISU	0.408807
32	KitchenQual	0.406756
12	BsmtCond	0.349831
31	KitchenAbvGr	0.347333
30	BedroomAbvGr	0.280257
2	LotShape	0.206243
19	TotalBsmtSF	0.173706
69	Neighborhood_MeadowV	0.173390
13	BsmtExposure	0.155340

Best Ridge

	Feature	Coef
46	PoolArea	0.545684
32	KitchenQual	0.387945
77	Neighborhood_SWISU	0.366843
31	KitchenAbvGr	0.325913
30	BedroomAbvGr	0.272939
12	BsmtCond	0.207743
2	LotShape	0.192425
39	GarageQual	0.189081
69	Neighborhood_MeadowV	0.183787
60	Neighborhood_Blueste	0.182990

Doubled Ridge

	Feature	Coef
46	PoolArea	0.685051
32	KitchenQual	0.420505
77	Neighborhood_SWISU	0.381683
31	KitchenAbvGr	0.350222
39	GarageQual	0.314452
30	BedroomAbvGr	0.286389
76	Neighborhood_OldTown	0.263192
60	Neighborhood_Blueste	0.230278
75	Neighborhood_NridgHt	0.219865
14	BsmtFinType1	0.214384

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 optimal values of lambda for ridge and lasso regression are 1.7 and 0.001 respectively. And R2 Score is also the same for both which is 0.88 (For both train and test).
We will use Lasso as it helps better in feature elimination, hence it is the final model.

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?

Our five most important predictor variables in the lasso model were-

- PoolArea
- Neighborhood_SWISU
- KitchenQual
- BstmCond
- KitchenAbvGr

After building another lasso model and removing these 5 features. Now we have the follow top five important predictor variables as -

- PoolArea
- Neighborhood_SWISU
- KitchenQual
- BedrooomAbvGr
- KitchenAbvGr

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?

A good model is robust and generalisable; it provides fewer predictions on test data. It won't be requiring huge data to be trained and will be more generalized with low variance and high bias. The problem with complex models is that it overfits.
The simple model shouldn't be simple enough to not learn anything, however it should not be overfitting and underfitting at the same time. The complexity of simple models can also be reduced using regularization and the bias- variance tradeoff has to be good enough.