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?

Answer:

The optimal value of alpha for lasso and ridge are 0.0001 and 0.7 respectively.
With increase of alpha accuracy decrease. It might increase a bit till the optimal hyper parameter value but the accuracy will decrease with the increase in alpha and model will become more parse.

The most important predictor variables before and after the change is implemented is as follows:

Features that lasso provides at optimal alpha (0.0001):

	Featuere	Coef
43	PoolArea	1.785626
67	Neighborhood_Gilbert	0.763425
68	Neighborhood_IDOTRR	0.516809
41	3SsnPorch	0.398264
26	FullBath	0.302911
65	Neighborhood_Crawfor	0.260334
40	EnclosedPorch	0.246211
9	BsmtQual	0.230478
53	LandContour_HLS	0.223526
1	LotArea	0.209502

Features that lasso provides at double the optimal alpha (0.0002):

	Featuere	Coef
43	PoolArea	1.750745
67	Neighborhood_Gilbert	0.632166
68	Neighborhood_IDOTRR	0.405947
26	FullBath	0.315832
41	3SsnPorch	0.267934
9	BsmtQual	0.219450
1	LotArea	0.213728
16	BsmtUnfSF	0.208574
65	Neighborhood_Crawfor	0.203769
53	LandContour_HLS	0.191694

Features that Ridge provides at optimal alpha (0.7):

	Feaure	Coef
43	PoolArea	1.530009
67	Neighborhood_Gilbert	0.567302
68	Neighborhood_IDOTRR	0.352070
26	FullBath	0.309764
41	3SsnPorch	0.280779
53	LandContour_HLS	0.241421
1	LotArea	0.212804
16	BsmtUnfSF	0.210351
65	Neighborhood_Crawfor	0.208491
54	LandContour_Low	0.198578

Features that Ridge provides at double the optimal alpha (1.4):

	Feaure	Coef
43	PoolArea	1.316088
67	Neighborhood_Gilbert	0.440710
26	FullBath	0.318515
68	Neighborhood_IDOTRR	0.244097
53	LandContour_HLS	0.218743
1	LotArea	0.218403
16	BsmtUnfSF	0.208221
41	3SsnPorch	0.190123
14	BsmtFinType2	0.181083
11	BsmtExposure	0.178149

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?

Answer:

r2_scores of both are almost same so we are using but lasso use more data set and better in feature elimation

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?

The five values that best describe the final model are as follows:

PoolArea ,
Neighborhood_Gilbert ,
Neighborhood_IDOTRR ,
FullBath,
3SsnPorch

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:

We should ensure that the model is robust and generalisable by regularizing the model and using a regularisation term with the RSS because the hyper parameter will ensure to strike the right balance between the model being too simple or too complex. Making the model more generalisable may take a toll on accuracy upto some extent but we can also have a look at the precision and recall of the model because sensitivity and specificity also play an important role in the model evaluation criteria. Together if all three are above average we may accept the model.
A very accurate model may have a chance of getting overfitted.