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 0.9 respectively.
If we increase the alpha(hyper parameter value) the accuracy of the model starts dropping gradually. 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):

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

Out[48]:

	Featuere	Coef
44	MiscVal	1.843010
40	EnclosedPorch	0.608121
25	BsmtHalfBath	0.593294
27	HalfBath	0.454099
22	LowQualFinSF	0.433486
67	Neighborhood_Gilbert	0.415242
68	Neighborhood_IDOTRR	0.411271
24	BsmtFullBath	0.394512
19	CentralAir	0.324535
23	GrLivArea	0.305384

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

	Featuere	Coef
48	YrSold_Old	1.697017
31	TotRmsAbvGrd	0.541710
33	GarageFinish	0.412867
30	KitchenQual	0.327390
67	Neighborhood_Gilbert	0.326034
68	Neighborhood_IDOTRR	0.321861
11	BsmtExposure	0.293883
29	KitchenAbvGr	0.287118
21	2ndFlrSF	0.270080
44	MiscVal	0.270026

Features that Ridge provides
at optimal alpha (0.9):

	Feaure	Coef
48	YrSold_Old	1.479578
31	TotRmsAbvGrd	0.537216
33	GarageFinish	0.419822
67	Neighborhood_Gilbert	0.377627
30	KitchenQual	0.342210
68	Neighborhood_IDOTRR	0.304335
21	2ndFlrSF	0.292788
29	KitchenAbvGr	0.286926
44	MiscVal	0.281953
19	CentralAir	0.262592

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

Out[168]:

	Feaure	Coef
48	YrSold_Old	1.264596
31	TotRmsAbvGrd	0.517219
33	GarageFinish	0.389861
30	KitchenQual	0.342352
67	Neighborhood_Gilbert	0.304537
29	KitchenAbvGr	0.275239
21	2ndFlrSF	0.234680
68	Neighborhood_IDOTRR	0.212173
60	Neighborhood_Blueste	0.199163
19	CentralAir	0.198257

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:

After creating model in both Ridge and Lasso are having similar r2_scores but as
lasso will penalise more on the dataset and can also help in feature elimination.
Choosing Lasso for final model

Question 3

After building the model, you realized 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:

YrSold_Old

TotRmsAbvGrd

GarageFinish

KitchenQual

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 regularize the model and using a regularisation term with the RSS because the hyper parameter willensure to have right balance between the model being too simple or too complex (variance and bias). Making the model more general 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.