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):

Featuere	Coef
MiscVal	1.843010
EnclosedPorch	0.608121
BsmtHalfBath	0.593294
HalfBath	0.454099
LowQualFinSF	0.433486
Neighborhood_Gilbert	0.415242
Neighborhood_IDOTRR	0.411271
BsmtFullBath	0.394512
CentralAir	0.324535
GrLivArea	0.305384

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

Featuere	Coef
MiscVal	1.803180
BsmtHalfBath	0.587252
EnclosedPorch	0.473518
HalfBath	0.430106
LowQualFinSF	0.425581
BsmtFullBath	0.392827
Neighborhood_Gilbert	0.351567
BsmtCond	0.304295
Neighborhood_IDOTRR	0.301270
GrLivArea	0.278336

Features that Ridge provides at optimal alpha (0.9):

Feaure	Coef
MiscVal	1.482910
BsmtHalfBath	0.574493
HalfBath	0.421184
LowQualFinSF	0.419674
BsmtFullBath	0.399218
Neighborhood_Gilbert	0.338057
EnclosedPorch	0.335071
TotRmsAbvGrd	0.286967
GrLivArea	0.266793
Neighborhood_IDOTRR	0.261202

Features that Ridge provides at double the optimal alpha (1.8):

Feaure	Coef
MiscVal	1.224349
BsmtHalfBath	0.551467
LowQualFinSF	0.399191
BsmtFullBath	0.391911
HalfBath	0.382626
TotRmsAbvGrd	0.335736
Neighborhood_Gilbert	0.275359
GrLivArea	0.224120
1stFlrSF	0.216488
EnclosedPorch	0.213262

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 we can see that the r2_scores are almost same for both of them but as lasso will penalise more on the dataset and can also help in feature elimination. Therefore we are going to consider that as our 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?

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

MiscVal : $ Value of miscellaneous feature
BsmtHalfBath : Basement half bathrooms
LowQualFinSF : Low quality finished square feet (all floors)
BsmtFullBath : Basement full bathrooms
HalfBath : Half baths above grade

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.