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 value of alpha obtained for Ridge is 2 and for Lasso is 0.001. With alpha=2, the R2 score of train and test dataset for the Ridge model were 0.897 and 0.820 respectively. And with alpha=0.001, the R2 score of train and test dataset for the Lasso model were 0.887 and 0.838 respectively.

After doubling the alpha values in the Ridge and Lasso, the prediction accuracy remained approximately similar. With alpha=4, the R2 score of train and test dataset for the Ridge model were 0.891 and 0.837 respectively. And with alpha=0.002, the R2 score of train and test dataset for the Lasso model were 0.881 and 0.835 respectively.

No change in the co-efficient values are observed in Ridge model for different values of alpha. The new models are created and demonstrated in the Jupyter notebook. Screenshot illustrating changes in the co-efficient are attached below.

Ridge model with alpha = 2

	Feaure	Coef
5	OverallQual	0.403619
26	HalfBath	0.375168
58	Neighborhood_Crawfor	0.318587
2	LotArea	0.295469
68	Neighborhood_NridgHt	0.285271
21	2ndFlrSF	0.223859
3	LotShape	0.223585
4	LandSlope	0.192710
7	MasVnrArea	0.189509
51	LotConfig_FR3	0.183385

Ridge model with alpha = 4

	Feaure	Coef
5	OverallQual	0.403619
26	HalfBath	0.375168
58	Neighborhood_Crawfor	0.318587
2	LotArea	0.295469
68	Neighborhood_NridgHt	0.285271
21	2ndFlrSF	0.223859
3	LotShape	0.223585
4	LandSlope	0.192710
7	MasVnrArea	0.189509
51	LotConfig_FR3	0.183385

Small changes in the co-efficient values are observed in Lasso model for different values of alpha.

Lasso model with alpha = 0.001

	Featuere	Coef
26	HalfBath	0.452667
5	OverallQual	0.407371
68	Neighborhood_NridgHt	0.327528
2	LotArea	0.298118
21	2ndFlrSF	0.249647
58	Neighborhood_Crawfor	0.242762
3	LotShape	0.233999
4	LandSlope	0.192793
51	LotConfig_FR3	0.172513
1	LotFrontage	0.159077

Lasso model with alpha = 0.002

	Featuere	Coef
5	OverallQual	0.407322
26	HalfBath	0.379411
68	Neighborhood_NridgHt	0.337685
2	LotArea	0.305520
3	LotShape	0.250106
21	2ndFlrSF	0.245462
58	Neighborhood_Crawfor	0.219146
4	LandSlope	0.194935
1	LotFrontage	0.166329
7	MasVnrArea	0.143343

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 optimum lambda value in case of Ridge and Lasso is as follows:-

- Ridge – 2
- Lasso – 0.001

We observe that the r2_scores are almost same for both of them, its 0.82 for Ridge and 0.83 for Lasso. Further, Lasso will penalise more on the dataset and can also help in feature elimination (as the coefficient value of some of the features become zero). Therefore we are going to consider Lasso as our final model, since Lasso has edge over Ridge 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 now?

Following are the five most important predictor variables in the original lasso model

- HalfBath with coefficient 0.452667

- OverallQual with coefficient 0.407371
- Neighborhood_NridgHt with coefficient 0.327528
- LotArea with coefficient 0.298118
- 2ndFlrSF with coefficient 0.249647

The five most important predictor variables of another model excluding the five most important predictor variables would be

58	Neighborhood_Crawfor	0.242762
3	LotShape	0.233999
4	LandSlope	0.192793
51	LotConfig_FR3	0.172513
1	LotFrontage	0.159077

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?

As Per, Occam's Razor— given two models that show similar 'performance' in the finite training or test data, we should pick the one that makes fewer assumptions on the test data due to following reasons:-

- Simpler models are usually more 'generic' and are more widely applicable
- Simpler models require fewer training samples for effective training than the more complex ones and hence are easier to train.
- Simpler models are more robust.
- Complex models tend to change wildly with changes in the training data set
- Simple models have low variance, high bias and complex models have low bias, high variance
- Simpler models make more errors in the training set.
- Complex models lead to overfitting — they work very well for the training samples, fail miserably when applied to other test samples.

Therefore, to make the model more robust and generalizable, make the model simple but not simpler which will not be of any use. Regularization can be used to make the model simpler. Regularization helps to strike the delicate balance between keeping the model simple and not making it too naive to be of any use.

For regression, regularization involves adding a regularization term to the cost that adds up the absolute values or the squares of the parameters of the model. Also, Making a model simple leads to Bias-Variance Trade-off.

A complex model will need to change for every little change in the dataset and hence is very unstable and extremely sensitive to any changes in the training data. A simpler model that extracts out some pattern followed by the data points given is unlikely to change wildly even if more points are added or removed.

Bias quantifies how accurate is the model likely to be on test data. A complex model can do an accurate job prediction provided there is enough training data. Models that are too naïve, for e.g., one that gives same answer to all test inputs and makes no discrimination whatsoever has a very large bias as its expected error across all test inputs are very high. Variance refers to the degree of changes in the model itself with respect to changes in the training data. Thus accuracy of the model can be maintained by keeping the balance between Bias and Variance as it minimizes the total error as shown in the below graph.