Advance Regression Assignment 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?*

Answers – After building the model we found that out of 28 values we found the Ridge and Lasso value to be as bellow

a) Ridge regression (Lambda) – 5
b) Lasso Regression (Lambda) – 0.0001

Doubling the value of the lambda means higher regularization and if we will continue to increase it, we will reach a point where model will not learn any relationship in data and that will be the case of underfitting.

Hence we have to find the optimum value of the lambda by hit and trial method.

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 – Though the performance by Ridge regression was better, it's better to use Lasso as it assigns the value to insignificant feature making us to choose the predictive variables. It is recommended to use simple but a robust model for the purpose.

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?*

Answer - Five most important predictor variables are as follows :
GrLivArea, RoofMatl_WdShngl, OverallQual, Neighborhood_NoRidge, GarageCars

	Feature	Coeff
14	GrLivArea	0.338970
106	RoofMatl_WdShngl	0.123036
3	OverallQual	0.120857
60	Neighborhood_NoRidge	0.065602
21	GarageCars	0.053796
...	...	...
0	MSSubClass	-0.035662
152	BsmtQual_Gd	-0.037945

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 – Model should be simple and generic in order to serve the data set. We can see that via the bias-variance trade-off as the model if simple it is more bias but less variance and more generalizable and that will perform or have similar result for training and test data set. It is also important to have the balance between the bias and variance to avoid overfitting and underfitting.