Advanced Regression – Subjective Questions

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 in ridge and lasso regression is the value of lambda that can be applied as regularization strength to achieve good accuracy of the model and avoid overfitting/underfitting.

   When the Alpha value is increased, the penalty term shrinks the coefficient towards zero and tends to move the model towards underfitting. The top coefficient also tend to change with increase in Alpha value.

   In our case study below are the Alpha values of both Ridge & Lasso

   Ridge Alpha value – 3.0

	Linear	Ridge	Lasso
GrLivArea	6.991340e-01	0.281383	0.616163
OverallQual	3.457998e-01	0.256668	0.387414
OverallCond	3.287419e-01	0.205495	0.228584
Condition2_PosN	1.062149e+00	0.201193	0.000000
1stFlrSF	1.117693e-01	0.182548	0.000000

   With Alpha value – 6.0

	Linear	Ridge	Lasso
GrLivArea	6.991340e-01	0.243997	0.616163
OverallQual	3.457998e-01	0.217885	0.387414
1stFlrSF	1.117693e-01	0.180307	0.000000
OverallCond	3.287419e-01	0.170425	0.228584
TotalBsmtSF	2.987021e-01	0.141736	0.201845

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: I would be choosing Lasso as the initial data set has many variables, and considering the problem statement that is to identify only the important variables that would determine the price of the house, Lasso would make the right fit as it will simplify the model by making the irrelevant feature coeeficients to Zero.

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: GrLivArea, OverallQual, TotalBsmtSF, 1stFlrSF, GarageCars are the top five predictor variables.
   OverallCond, YearBuilt, GarageArea, LotArea, MSZoning_FV are the next important variables calculated after removing the above varaiables.

	Linear	Ridge	Lasso
BsmtFinSF1	4.181011e-01	0.281244	0.325554
BsmtUnfSF	3.969743e-01	0.242515	0.273741
2ndFlrSF	2.593438e-01	0.208577	0.225528
Condition2_PosN	1.121619e+00	0.205566	0.000000
TotRmsAbvGrd	2.613122e-01	0.197390	0.214992

4. How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?
   Answer: A model that performs well on unseen data and exhibits consistent performance across different datasets is considered more reliable and trustworthy. This can be achieved by Implementing Cross-validation techniques and testing the model with different sets of the data would be one way to check the robustness of the model. Identifying the most important features of the model also helps to avoid over fitting the model.