1. 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 LAMBDA we got in case of Ridge and Lasso is :
  a. Ridge = 0.4
  b. Lasso = 0.0001
- The changes in the model if we choose to double the value of alpha for both ridge and lasso are:

Model	Alpha	r2_score on Train	r2_score on Test	MSE on Train	MSE on Test
Ridge	0.4	0.8863	0.8882	0.00257	0.00321
	0.8	0.8861	0.8877	0.00257	0.00323
Lasso	0.0001	0.8860	0.8852	0.00258	0.0033
	0.0002	0.8849	0.8630	0.00260	0.0039

  a. When we double the alpha value in case of ridge there is a slight increase in the mean squared error whereas the r2_score slightly decreased on both train and test data.
  b. When we double the alpha value in case of lasso there is a slight increase in the mean squared error whereas the r2_score slightly decreased on train data and slightly increase on test data.
- Top 5 correlated features when alpha is 0.0002 in Lasso are:
  a. Total_sqr_footage
  b. OverallQual
  c. OverallCond
  d. BsmtUnfSF
  e. GrLivArea

- Top 5 correlated features when alpha is 0.8 in Ridge are:
  a. OverallQual
  b. Total_sqr_footage
  c. GrLivArea
  d. OverallCond
  e. TotalBsmtSF

2. 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 :-

- The optimum lambda value in case of Ridge and Lasso is as follows: -
  Ridge – 0.4
  Lasso – 0.0001
- The Mean Squared Error in case of Ridge and Lasso are:
  Ridge – 0.003
  Lasso – 0.0033
- The Mean Squared Error of both the models are almost same.
- Since Lasso helps in feature reduction (as the coefficient value of some of the features become zero), Lasso has a better edge over Ridge and should be used as the final model.

3. 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?

Answer :-

- OverallQual, Total_sqr_footage, GrLivArea, OverallCond, TotalBsmtSF are the five most important predictor variables in the lasso model. Hence we need to exclude these 5 variables from our model.
- Top 5 correlated features when alpha is 0.0001 in Lasso are:
  a. LotFrontage
  b. Fireplaces
  c. Total_porch_sf
  d. Total_Bathrooms
  e. GarageCars

4. Question 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 :-

- The model is expected to be as simple as possible and simpler models are considered as more ‘generic’, though its accuracy will be decreased but it will be more robust.
- This can be understood from the Bias-Variance trade-off. The simpler the model the more the bias but less variance becoming generalizable. Whereas the complex model will have high variance and low bias.
- Sometimes underfitting and overfitting are the problems associated with the model. Hence, it is important to have balance in Bias and Variance to avoid such problems. This is possible with “Regularization”.
- Regularization helps in managing the model complexity by essentially shrinking the coefficients towards zero. This avoids the model becoming too complex, thus reducing the risk of overfitting.
- Regularization method should be used to keep the model optimum simpler. It penalizes the model if it becomes more complex.

- Regularization method helps to achieve the Bias-Variance trade off. It compromises by increasing bias to a optimum position where Total Error is minimum.
- This point also known as Optimum Model Complexity where Model is sufficient simpler to be generalizable and also complex enough to be robust.
- Making a model simple lead to Bias-Variance trade off.