Problem Statement - Part II

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?

Ans : The ideal alpha value for ridge is 10.

The ideal alpha value for ridge is 100.

Once the double alpha for the lasso and ridge, or 20 and 200, is made Regarding Ridge: When alpha rises, Coeff values also rise. Additionally, the train data's r2_score drops from.807 to 0.45. Regarding Lasso: More features were eliminated from the model when the alpha value rose. However, the r2score in the test and train data is similarly lowered by 1%. Principal attributes: OverallQual, overallQual, Neighborhood_NoRidge, Neighborhood_NridgHt Veenkar Neighborhood

```
The most important predictor variables are as follows:
```

	Ridge Doubled Alpha Co-Efficient
CentralAir_Y	0.029878
Neighborhood_StoneBr	0.024774
Alley_Pave	0.021760
MSSubClass_70	0.019776
PavedDrive_P	0.016999
LandContour_HLS	0.016946
Condition1_Norm	0.016638
Neighborhood_Veenker	0.016433
KitchenQual_Ex	0.015628
BsmtQual_Ex	0.015514
MasVnrType_Stone	0.015092
SaleType_Con	0.015088
HouseStyle_1.5Unf	0.014682
Condition1_PosN	0.013992
PavedDrive_Y	0.013718
SaleType_Oth	0.013521
RoofMatl_WdShngl	0.012753
HouseStyle_2.5Unf	0.012678
SaleType_CWD	0.011919
SaleCondition_Partial	0.011791

	Lasso Co-Efficient
CentralAir_Y	0.032557
Neighborhood_StoneBr	0.025251
Alley_Pave	0.021622
MSSubClass_70	0.017581
Condition1_Norm	0.017280
LandContour_HLS	0.016686
BsmtQual_Ex	0.016053
Neighborhood_Veenker	0.015890
KitchenQual_Ex	0.015813
MasVnrType_Stone	0.014390
PavedDrive_P	0.014058
Condition1_PosN	0.013030
PavedDrive_Y	0.012495
SaleCondition_Partial	0.011187
BsmtCond_TA	0.010814
HouseStyle_2.5Unf	0.010779
LotConfig_CulDSac	0.009210
HouseStyle_1.5Unf	0.009027
OverallCond	0.007399
TotRmsAbvGrd	0.007013

	Ridge Co-Efficient
CentralAir_Y	0.029878
Neighborhood_StoneBr	0.024774
Alley_Pave	0.021760
MSSubClass_70	0.019776
PavedDrive_P	0.016999
LandContour_HLS	0.016946
Condition1_Norm	0.016638
Neighborhood_Veenker	0.016433
KitchenQual_Ex	0.015628
BsmtQual_Ex	0.015514
MasVnrType_Stone	0.015092
SaleType_Con	0.015088
HouseStyle_1.5Unf	0.014682
Condition1_PosN	0.013992
PavedDrive_Y	0.013718
SaleType_Oth	0.013521
RoofMatl_WdShngl	0.012753
HouseStyle_2.5Unf	0.012678
SaleType_CWD	0.011919
SaleCondition_Partial	0.011791

	Lasso Doubled Alpha Co-Efficient
CentralAir_Y	0.031768
Neighborhood_StoneBr	0.018404
BsmtQual_Ex	0.017478
Alley_Pave	0.016133
KitchenQual_Ex	0.014219
Condition1_Norm	0.013853
LandContour_HLS	0.013637
MasVnrType_Stone	0.012956
MSSubClass_70	0.011707
SaleCondition_Partial	0.011525
BsmtCond_TA	0.010600
LotConfig_CulDSac	0.008313
PavedDrive_Y	0.008285
OverallCond	0.007504
TotRmsAbvGrd	0.007125
BldgType_TwnhsE	0.006576
PavedDrive_P	0.006488
MasVnrType_BrkFace	0.006029
ExterCond_TA	0.004550
RoofStyle_Hip	0.002331

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?

Ans: Since Lasso offers the opportunity to pick features as well, we shall go with it. Without compromising the precision of the model, it has eliminated undesired aspects. It results in a simple, accurate, and generalized model.

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?

Ans : Neighborhood_NoRidge, Neighborhood_NridgHt, 2ndFlrSF, OverallQual, and Neighborhood_Veenker are the top 5 features. Following their removal, the model's accuracy dropped from 81% and 80% to 58% and 55%. Currently, the most desirable features are: Top 5 traits following the removal of 5 primary predictors FirstFlrSF, TotalBsmtSF, MSSubClass_90, MSSubClass_120, and HouseStyle_1Story

]:

	Lasso Co-Efficient
HouseStyle_2.5Unf	0.062613
HouseStyle_2.5Fin	0.049470
Neighborhood_StoneBr	0.043728
Neighborhood_Veenker	0.041480
CentralAir_Y	0.041457

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?

Ans : Three characteristics are needed for the model to be resilient and generalizable:

1. Model accuracy should be greater than 70–75%; in our instance, it is coming in at 80% for train and 81% for test, which is accurate.
2. Every feature's P-value is less than 0.05.
3. All characteristics have a VIF of less than 5.

As a result, we are certain of the model's robustness and generalizability.