1.Doubling Hyperparameters:

Before Doubling:

OverallQual	-2.186634
Condition2_RRAn	-2.723929
MSZoning_RM	0.741685
MSZoning_RH	0.650330
MSZoning_RL	0.585064

After doubling the hyper parameters, we can see the coefficients have been reduced. As we have given more emphasis on regularization.

BsmtUnfSF	-1.603872
MSSubClass	-1.031653
LotArea	0.359718
OverallQual	0.338057
MSZoning_RH	0.296470

2.The optimal Value for Ridge is 0.6 and Lasso is 0.1

The r2 score for Ridge is:

```
0.8849946265977674
0.8301860302300867
```

The R2 score for Lasso is

```
0.8708588652221473
0.8418204702135157
```

Both are performing almost same. **But when they both perform same I would pick Lasso because it would give me feature selection as well.**

Ridge is good in scenarios of Overfitting because it penalizes more.

Ridge regression adds "*squared magnitude*" of coefficient as penalty term to the loss function. Here the *highlighted* part represents L2 regularization element.

$$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

Lasso Regression (Least Absolute Shrinkage and Selection Operator) adds "*absolute value of magnitude*" of coefficient as penalty term to the loss function.

$$\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

3.After removing top 5 and built the model with optimal hyper parameter

The top 5 variables are

8	Neighborhood_StoneBr	-0.755567
0	YearBuilt	-1.985727
14	RoofStyle_Shed	-3.063636
21	Exterior1st_BrkComm	0.968195
19	RoofMatl_WdShake	0.841667

4. **Ridge regression** adds "*squared magnitude*" of coefficient as penalty term to the loss function. Here the *highlighted* part represents L2 regularization element.

$$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

Lasso Regression (Least Absolute Shrinkage and Selection Operator) adds "*absolute value of magnitude*" of coefficient as penalty term to the loss function.

$$\sum_{i=1}^{n}(Y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

From the above equations we can see that Lambda parameter regularizes the model

If Lambda increases then regularization increases making the model more simple and generalizable. Decreasing the variance and increasing the bias.

If Lambda is 0 the model would become a Linear regression model with no regularization which can increase the complexity.

We must choose Lambda in such a way that there is bias variance tradeoff.

If Lambda increases the accuracy decreases because we are making the model simple by regularizing it.

When Lambda increases the coefficients decrease.R2 score of the model decreases Which can cause under fitting.

There are some ways to regularize a model

1.Model Selection criteria.

Residual sum of Squares

Total sum of squares.

If ^yi is the predicted value and yi is the actual value, then **RSS** is given as follows:

∑ni=1(^yi−y)2

(^yi−y) is often called the **residue.** Also, the total sum of squares TSS is written as the sum of the explained sum of squares (ESS) and RSS. : TSS=ESS+RSS

We can use Adjusted R2 for model selection:

Adjusted R2=1−(RSS/(n−d−1))/(TSS/(n−1))

2.Feature Selection.

RFE- Recursive Feature Elimination