

Q1: 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: Optimal best hyperparameter alpha for Ridge is 2 and Lasso is 0.001 in my regression model. When we double the alpha * 2 and rebuilt with Ridge and Lasso model, there will a change in the R2 score reduction. Top predictor influence sales price positively or negatively are given below.

Result in Optimal value of alpha

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.92	0.91	0.91
1	R2 Score (Test)	0.91	0.91	0.90

Result after, Double alpha i.e., alpha * 2

Ridge

r2_score_train: 0.9057615236671366

r2_score_test: 0.9017300712996023

Lasso

r2_score_train: 0.9057615236671366

r2_score_test: 0.9017300712996023

Important Predictor using double alpha

```
Ridge Positive Coef
OverallQual_10    1.86
OverallQual_9     1.42
OverallQual_8     1.13
OverallQual_7     0.82
OverallQual_6     0.70
Name: Ridge, dtype: float64
```

```
Lasso Positive Coef
OverallQual_10    1.08
OverallQual_9     0.74
OverallQual_8     0.42
Neighborhood_Crawfor 0.38
GrLivArea         0.35
Name: Lasso, dtype: float64
```

```
Ridge Negative Coef
BedroomAbvGr_8    -0.95
Exterior1st_Others -0.38
Neighborhood_Mitchel -0.37
BedroomAbvGr_6    -0.36
KitchenQual_Fa     -0.35
Name: Ridge, dtype: float64
```

```
Lasso Negative Coef
KitchenQual_Fa    -0.26
KitchenQual_TA    -0.19
BsmtQual_Gd       -0.16
Age               -0.15
KitchenQual_Gd    -0.13
Name: Lasso, dtype: float64
```

Q2: 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: Lasso regression would be a better choice as it shrinks coefficients toward Zero and performs variable selection which helps in feature elimination for a robust model.

Q3: 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: Top 5 variables in the current model:

1. OverallQual_10 1.08
2. OverallQual_9 0.74
3. OverallQual_8 0.42
4. Neighborhood_Crawfor 0.38
5. GrLivArea 0.35

Result after removing the Top 5 predictor

1. TotRmsAbvGrd_Others 0.30
2. 1stFlrSF 0.26
3. GarageType_BuiltIn 0.26
4. Exterior1st_BrkFace 0.24
5. FullBath_3 0.22

Q4: 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: A model needs to be made robust and generalizable so that they are not impacted by outliers in the training data. The model should also be generalisable so that the test accuracy is not lesser than the training score. The model should be accurate for datasets other than the ones which were used during training. Too much weightage should not be given to the outliers so that the accuracy predicted by the model is high. To ensure that this is not the case, the outlier analysis needs to be done and only those which are relevant to the dataset need to be retained. Those outliers which it does not make sense to keep must be removed from the dataset. This would help increase the accuracy of the predictions made by the model. Confidence intervals can be used (typically 3-5 standard deviations). This would help standardize the predictions made by the model. If the model is not robust, it cannot be trusted for predictive analysis.