

**Problem Statement - Part II****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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer:

Ridge and Lasso are powerful techniques in linear regression that play a significant role in simplifying complex models. These techniques achieve this by introducing regularization, which essentially means they impose constraints on the model parameters, thereby helping in reducing the complexity of the model by reducing the influence of the insignificant features.

Ridge and Lasso Regression:

Both Ridge and Lasso are commonly used methods for predicting a continuous target variable based on one or more predictor variables. However, linear regression can be prone to overfitting, especially when dealing with datasets with multicollinearity, where predictor variables are highly correlated.

- **Lasso (Least Absolute Shrinkage and Selection Operator):** Lasso addresses the overfitting issue by adding a penalty term to the loss function that the linear regression model tries to minimize. This penalty minimizes some of the coefficients to become exactly zero, effectively performing feature selection. In simpler terms, Lasso identifies and eliminates irrelevant features, simplifying the model's structure.
- **Ridge:** In contrast, Ridge introduces a penalty term, which is the square of the coefficients, into the loss function. This penalty encourages all coefficients to be small but not exactly zero. While Ridge may not completely eliminate features, it significantly reduces their impact on the model's predictions. Ridge aims to strike a balance by making all features contribute somewhat equally while preventing any single feature from dominating the predictions.

Optimal Alpha and the Trade-Off:

In both Ridge and Lasso regression, the strength of regularization is controlled by a hyperparameter known as Alpha. A key task in implementing these techniques is to determine the optimal value of Alpha. This is typically achieved through cross-validation techniques such as GridSearchCV, which is readily available in the library SciKit-Learn.

**Problem Statement - Part II**

The optimal Alpha value is crucial because it represents a delicate trade-off between two key aspects of model performance: bias and variance. A low Alpha allows the model to fit the training data closely, resulting in low bias but high variance. Conversely, a high Alpha imposes stronger regularization, reducing the variance but potentially introducing bias. Therefore, finding the right balance is essential to ensure that the model generalizes well to unseen data.

**Doubling Alpha and Its Impact:**

Now, let's consider the scenario where we choose to double the value of Alpha in both Ridge and Lasso regression. This decision amplifies the regularization effect, leading to a simpler model.

- Ridge Regression: With a higher Alpha, Ridge pushes the coefficients further towards zero. Although they may not reach exactly zero, they become exceedingly small. This means that the influence of certain features on the model's predictions is drastically reduced. Ridge ensures that all features remain in the model but with significantly diminished importance.
- Lasso Regression: In Lasso, doubling Alpha intensifies the feature selection aspect. More coefficients are driven all the way to zero, effectively removing the corresponding features from the model.

Doubling Alpha strengthens regularization, leading to simpler models by reducing the impact of features (Ridge) or eliminating irrelevant features altogether (Lasso).

**Question 2**

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one would you choose to apply to and why?

Answer: As per the model created for the Advanced regression assignment, the metrics were as below:

**Problem Statement - Part II**

Ridge Regression:

'alpha': 10.0

Lasso Regression:

Lasso(alpha=0.001)

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.945544	0.934207	0.925014
1	R2 Score (Test)	0.877366	0.909501	0.911325
2	RSS (Train)	6.928168	8.370543	9.540062
3	RSS (Test)	6.545967	4.830666	4.733306
4	MSE (Train)	0.082618	0.090812	0.096949
5	MSE (Test)	0.122530	0.105259	0.104193

Ridge: It will be chosen if all features are required and more manual control on feature selection is necessary and multicollinearity is present

Lasso: It will be chosen when only the most important features are required for prediction and less manual intervention is required. Also, if the model is to be made simple

**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?

**Problem Statement - Part II****Answer:**

After building the model the top 5 features with highest coefficients are as below

Neighborhood_Crawfor	0.087922
Exterior1st_BrkFace	0.056177
OverallQual	0.052759
Functional	0.052490
Neighborhood_BrkSide	0.044368

If these are not available in the dataset, then after dropping the above Features from the  $X_{\text{test}}$ ,  $X_{\text{Train}}$ ,  $y_{\text{test}}$  and  $y_{\text{train}}$ , will redo the Lasso Regression and come up with new set of top features.

After redoing Lasso Regression with Reduced set we get the below metrics

	<b>Metric</b>	<b>Lasso Regression (Reduced Set)</b>
<b>0</b>	R2 Score (Train)	0.912454
<b>1</b>	R2 Score (Test)	0.893769
<b>2</b>	RSS (Train)	11.138075
<b>3</b>	RSS (Test)	5.670413
<b>4</b>	MSE (Train)	0.104754
<b>5</b>	MSE (Test)	0.114042

And the top 5 Features this time are

Neighborhood_Somerst	0.061600
BsmtCond	0.060003
Neighborhood_NridgHt	0.052846
OverallCond	0.050285
Neighborhood_StoneBr	0.038978

**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?

**Problem Statement - Part II**

Answer:

A model can be considered as robust when:

- It performs well on both Train and test data
- Its performance metrics like accuracy or mean-Squared-error do not swing based on different datasets or subsets
- Outliers and other noisy data don't affect the model's accuracy

Also, when Accuracy is considered, to make a robust model we have to increase the Bias, which in turn will reduce the variance and so will the Accuracy.