LASSO REGRESSION

&

RIDGE REGRESSION

@DataByteSun

Lasso Regression

A Statistical Approach for Predicting Outcomes

@DataByteSun

What is Lasso Regression?

- **Definition:** Lasso regression is a linear regression technique that includes an L1 regularization (penalty) term, which minimizes the sum of the squared residuals (RSS) and the absolute values of the coefficients. This penalty encourages sparsity by shrinking some coefficients to zero, effectively performing feature selection and simplifying the model.

- **Goal:** The main goals of lasso regression are to reduce overfitting by preventing the model from becoming too complex, and to automatically select the most important features by setting irrelevant coefficients to zero, making the model more interpretable and efficient.

Key Concepts in Lasso Regression

- **L1 Regularization:** Lasso adds a penalty proportional to the absolute value of the coefficients to the linear regression model, encouraging sparsity (zeroing out some coefficients).
- **Feature Selection:** Lasso helps in automatic feature selection by shrinking less important feature coefficients to zero.
- **Overfitting Prevention:** The regularization term in Lasso prevents overfitting by discouraging overly complex models.
- **Shrinkage:** Lasso reduces the magnitude of model coefficients, making the model more generalizable.
- **Tuning Hyperparameter (Alpha):** The strength of the regularization is controlled by the hyperparameter alpha. A higher alpha increases regularization, leading to more coefficients being set to zero.
- **Better Performance with High-Dimensional Data:** Particularly effective when dealing with datasets with many features ($p > n$), as it reduces model complexity.

Assumptions of Lasso Regression

- **Linearity:** Relationship between x and y is linear.
- **Independence:** The residuals (errors) should be independent of each other.
- **Homoscedasticity:** Constant variance of errors.
- **Normality:** Errors are normally distributed.
- **Multicollinearity (optional):** requires that independent variables have no or low to moderate correlations with one another.
- **No Outliers:** Lasso is sensitive to outliers.
- **Sufficient Data:** Ensure that you have enough data for reliable coefficient estimation.
- **Tuning the Regularization Parameter:** α should be appropriately chosen using cross-validation.

Lasso Regression Equation

The objective of lasso regression is to minimize the following cost function:

$$\text{Cost Function} = \text{RSS} + \lambda\ |\beta i|$$

Where,

- RSS (Residual Sum of Squares) is the sum of squared errors (the usual loss function for linear regression).

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Where,

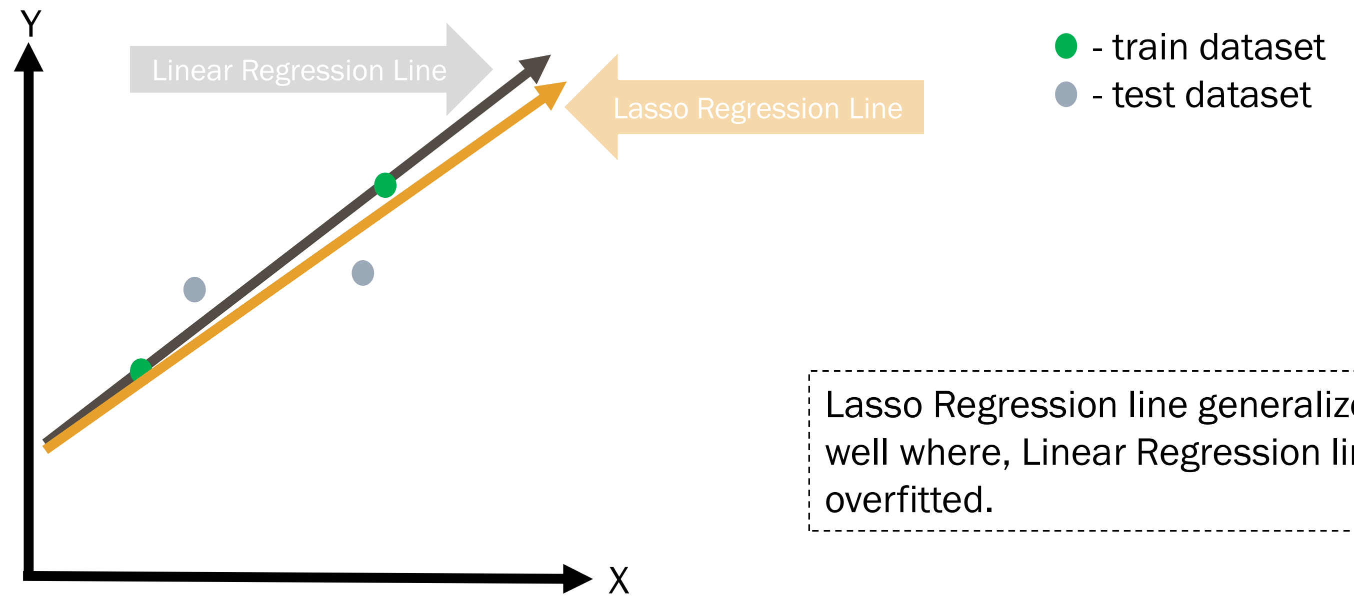
- y_i is the true value for the i-th observation.
- $\hat{y}_i$ is the predicted value for the i-th observation.

- λ (lambda) is the regularization parameter (also called the penalty term).
- βi represents the coefficients of the predictors.

How Lasso Regression Works

- **Linear Model:** Lasso fits a linear model to the data by finding the line (or hyperplane in higher dimensions) that minimizes the error between the predicted and actual values.
- **Penalty:** The regularization term (penalty) prevents the model from fitting too closely to the data (which would lead to overfitting).
- **Shrinkage:** Some feature coefficients will be shrunk to exactly zero, effectively removing those features from the model.

Lasso Regression



Lasso Regression line generalize well where, Linear Regression line overfitted.

@DataByteSun

When to Use Lasso Regression

- **When you have a large number of features:** Lasso regression is particularly useful when you have more predictors than observations or when there are many correlated predictors. It can help by automatically reducing the number of features used in the model.
- **When you want to perform feature selection:** Lasso's ability to shrink coefficients to zero means it's a useful tool for performing automatic feature selection, simplifying the model and focusing on the most important variables.
- **When there is multicollinearity:** In the presence of highly correlated features, lasso can help by selecting one variable and shrinking others to zero, effectively handling multicollinearity.

Ridge Regression

A Statistical Approach for Predicting Outcomes

@DataByteSun

What is Ridge Regression?

- **Definition:** Ridge Regression is a type of linear regression that adds an L2 penalty (the square of the coefficients) to the model, helping to prevent overfitting by shrinking the coefficients.
- **Goal:** The goal is to improve model stability and accuracy by limiting the size of the coefficients, using the L2 regularization term.

Key Concepts in Ridge Regression

- **L2 Regularization:** Ridge regression adds a penalty proportional to the sum of the squared coefficients to the cost function.
- **Multicollinearity Handling:** Ridge regression is particularly useful when the features are highly correlated (multicollinearity), as it reduces the impact of correlated predictors by shrinking their coefficients.
- **Shrinkage:** Ridge regression reduces the magnitude of coefficients, but unlike Lasso Regression, it does not set them to zero. This means all features remain in the model, but with smaller values.
- **Tuning Hyperparameter (Alpha):** The strength of the regularization is controlled by the hyperparameter alpha. A higher alpha increases regularization, leading to more coefficients being near to zero.

Assumptions of Ridge Regression

- **Linearity:** Relationship between x and y is linear.
- **Independence:** The residuals (errors) should be independent of each other.
- **Homoscedasticity:** Constant variance of errors.
- **Normality:** Errors are normally distributed.
- **Multicollinearity (optional):** requires that independent variables have no or low to moderate correlations with one another.
- **Sufficient Data:** Ensure that you have enough data for reliable coefficient estimation.
- **Tuning the Regularization Parameter:** α should be appropriately chosen using cross-validation.

Ridge Regression Equation

The objective of lasso regression is to minimize the following cost function:

$$\text{Cost Function} = \text{RSS} + \lambda\ (\beta i)^2$$

Where,

- RSS (Residual Sum of Squares) is the sum of squared errors (the usual loss function for linear regression).

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Where,

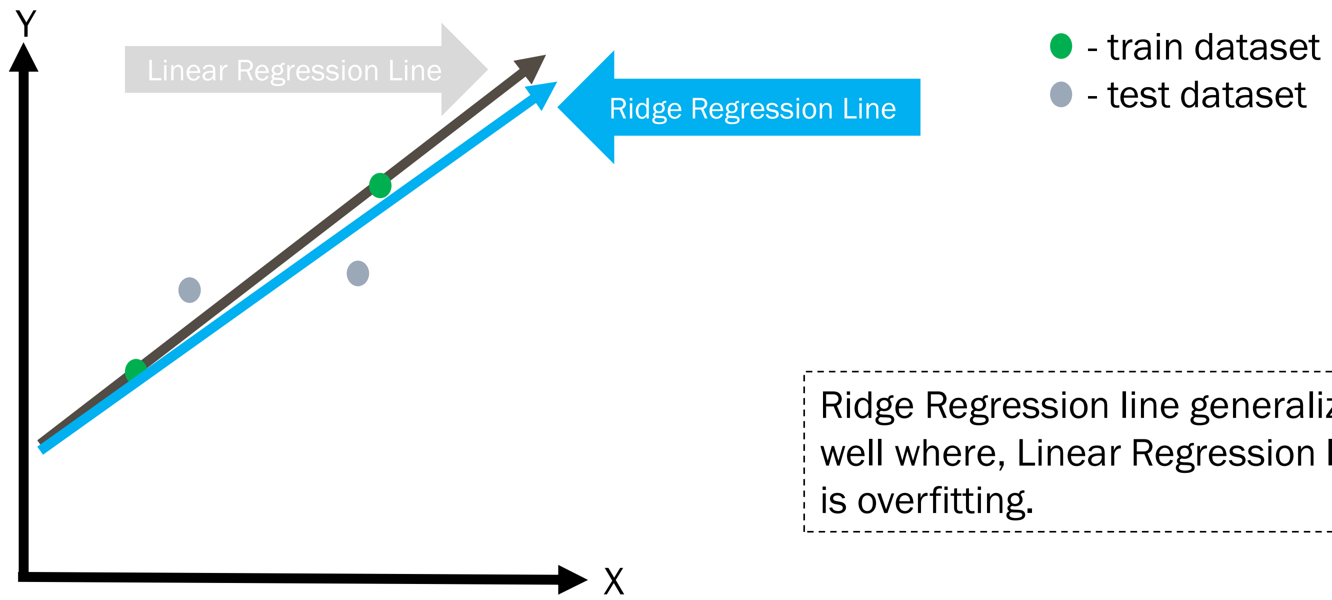
- y_i is the true value for the i-th observation.
- $\hat{y}_i$ is the predicted value for the i-th observation.

- λ (lambda) is the regularization parameter (also called the penalty term).
- βi represents the coefficients of the predictors.

How Ridge Regression Works

- **Linear Model:** Lasso fits a linear model to the data by finding the line (or hyperplane in higher dimensions) that minimizes the error between the predicted and actual values.
- **Penalty:** The regularization term (penalty) prevents the model from fitting too closely to the data (which would lead to overfitting).
- **Shrinkage:** Some feature coefficients will be shrunk near to zero, effectively handling those features for the model.

Ridge Regression



Ridge Regression line generalize well where, Linear Regression line is overfitting.

@DataByteSun

When to Use Ridge Regression

- **Many Features, All Important:** Use Ridge when you believe all features contribute to the model and don't need feature selection.
- **Multicollinearity:** Use Ridge if features are highly correlated to stabilize the model.
- **Predictive Power:** Use Ridge for better predictive accuracy when you don't want to remove features.
- **Feature Selection Not Needed:** Use Ridge if you don't need to remove irrelevant features.
- **Avoid Overfitting, Keep All Features:** Ridge helps reduce overfitting by shrinking coefficients but retains all features.
- **More Features than Data Points:** Ridge works well when more features than data points.

Lasso Regression VS Ridge Regression

Lasso Regression

- **Lasso** uses the **L1 regularization** (absolute value of coefficients), which can shrink some coefficients to zero, thus performing **feature selection.**
- Use when only some of features are important.
- Lasso Regression can shrink the slope **all the way to 0.**

Cost Function = RSS +λ $|\beta i|$

Ridge Regression

- **Ridge** uses **L2 regularization** (square of coefficients), which tends to shrink coefficients but doesn't set them exactly to zero. Ridge regularization **doesn't perform feature selection** in the same way lasso does.
- Use when most or all features are useful.
- Ridge Regression can only shrink the slope asymptotically **close to 0.**

Cost Function = RSS +λ $(\beta i)^2$

@DataByteSun

Thank You

\- @DataByteSun