Ridge Regression & Lasso Regression

DATA SCIENCE WITH PYTHON

What is Regularization

Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting. Regularization on an over-fitted model

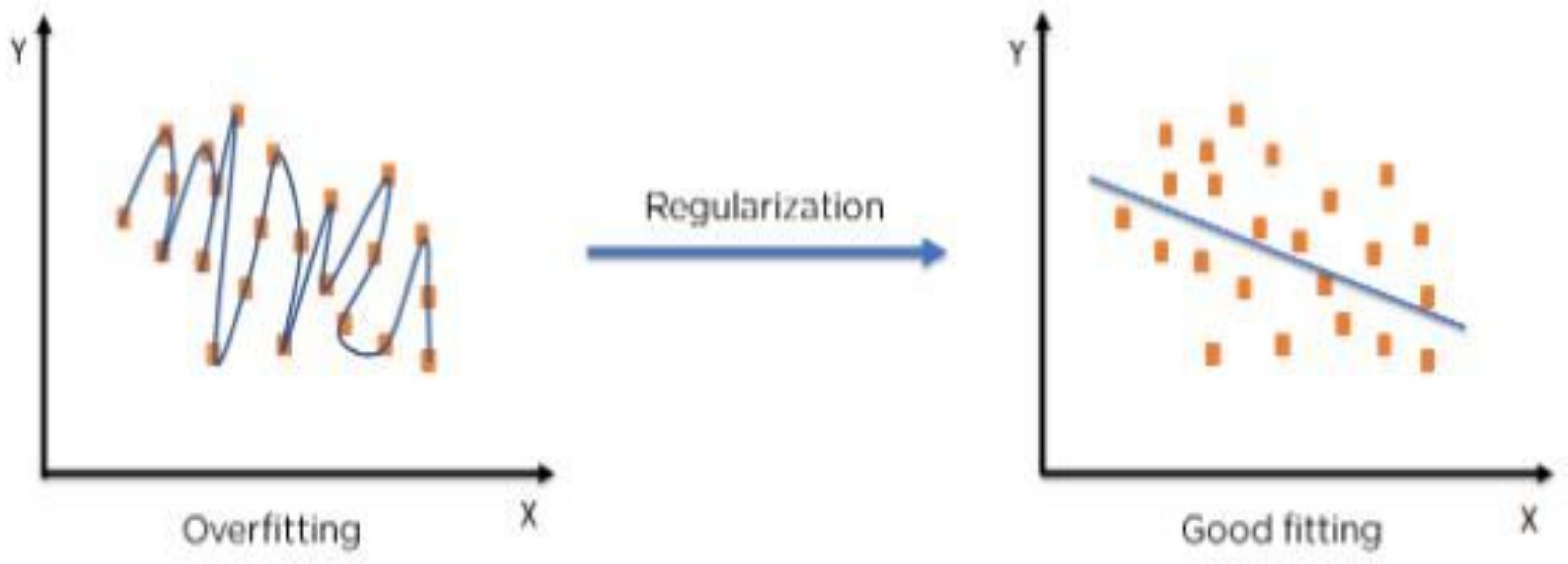
Using Regularization, we can fit

our machine learning model

appropriately on a given test

set and hence reduce the errors

in it.

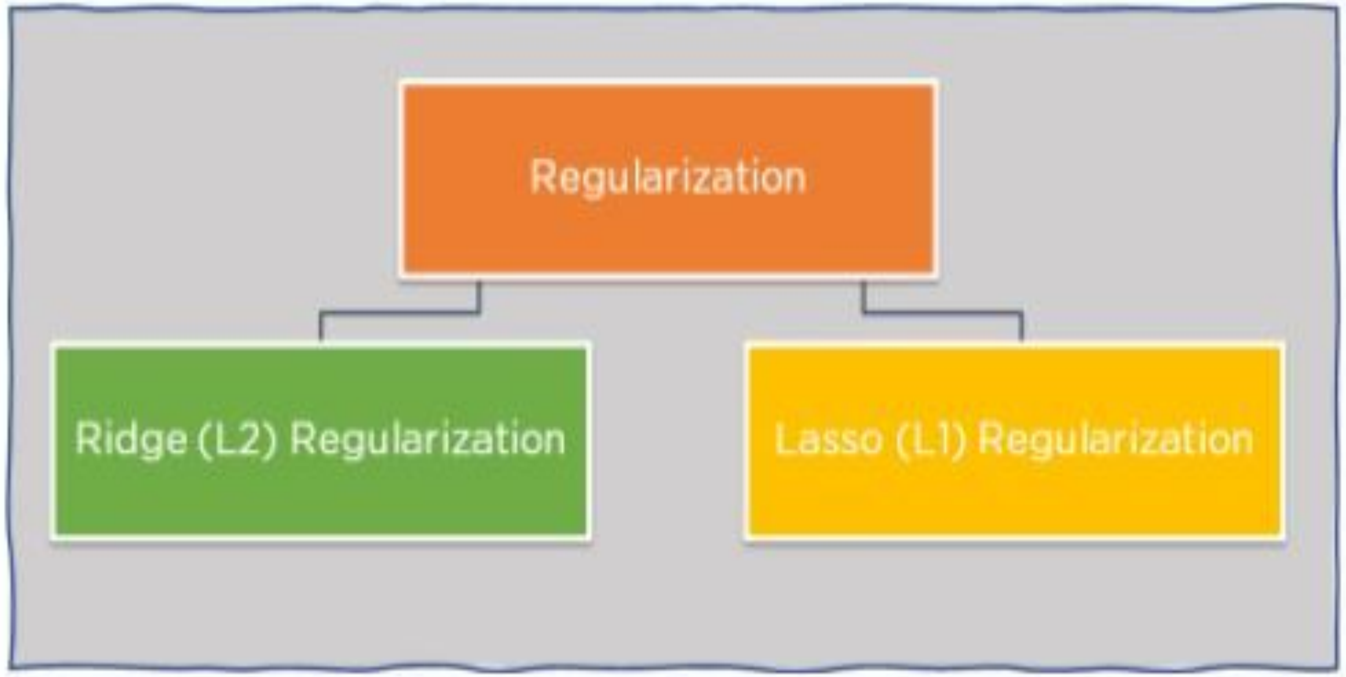


Regularization Techniques

There are two main types of regularization techniques:

1.Ridge Regularization and

2.Lasso Regularization.



Both Ridge and Lasso regression are types of linear regression with regularization techniques that help prevent overfitting.

Ridge Regularization

Also known as Ridge Regression, it modifies the over-fitted or under fitted models by adding the penalty equivalent to the sum of the squares of the magnitude of coefficients.

This means that the mathematical function representing our machine learning model is minimized and coefficients are calculated. The magnitude of coefficients is squared and added.

Ridge Regression performs regularization by shrinking the coefficients present. The function depicted below shows the cost function of ridge regression :

What is Ridge Regression or (L2 Regularization) Method?

Ridge regression, also known as L2 regularization, is a technique used in linear regression to prevent overfitting by adding a penalty term to the loss function. This penalty is proportional to the square of the magnitude of the coefficients (weights).

Ridge Regression is a version of linear regression that includes a penalty to prevent the model from overfitting, especially when there are many predictors or not enough data.

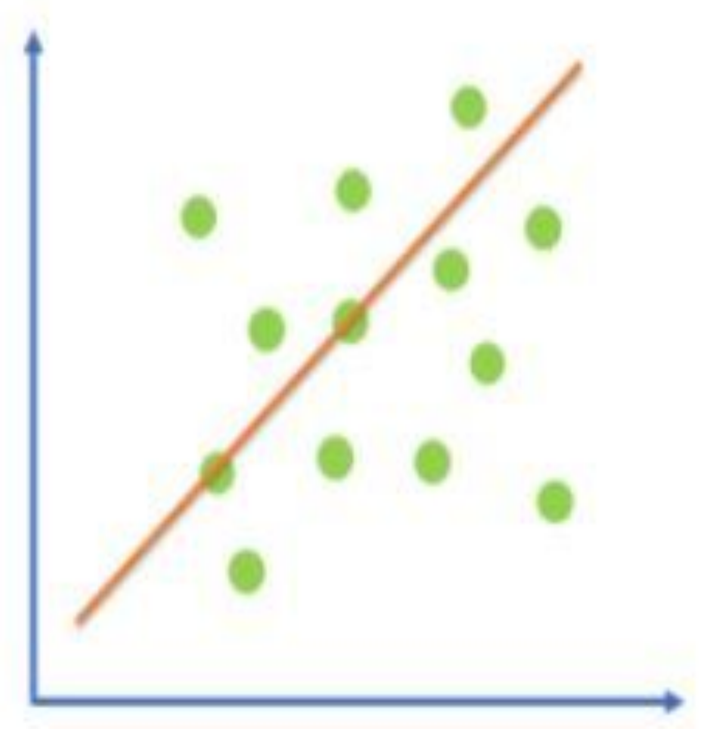
The standard loss function (mean squared error) is modified to include a regularization term:

$$\text{Loss} = \text{MSE}+\lambda \sum_{i=1}^{n} wi^2$$

Here, λ is the regularization parameter that controls the strength of the penalty, and wi are the coefficients

Cost Function of Ridge Regression

$$Loss=\sum(yi-\hat{y}i)^2+\lambda\sum wj^2$$



where:

- (yi−y^i)2(y_i - \hat{y}_i)^2(yi−y^i)2 is the Mean Squared Error (MSE).
- λ∑wj2\lambda \sum w_j^2λ∑wj2 is the L2 regularization term (sum of squared weights).
- λ\lambdaλ is the regularization strength (higher λ\lambdaλ shrinks coefficients more).

Cont..

In the cost function, the penalty term is represented by Lambda λ. By changing the values of the penalty function, we are controlling the penalty term.

The higher the penalty, it reduces the magnitude of coefficients. It shrinks the parameters.

Therefore, it is used to prevent multicollinearity, and it reduces the model complexity by coefficient shrinkage.

Effect:

Reduces overfitting by shrinking coefficients **but does not eliminate them** (coefficients become smaller but not zero).

Works well when **many features** are relevant.

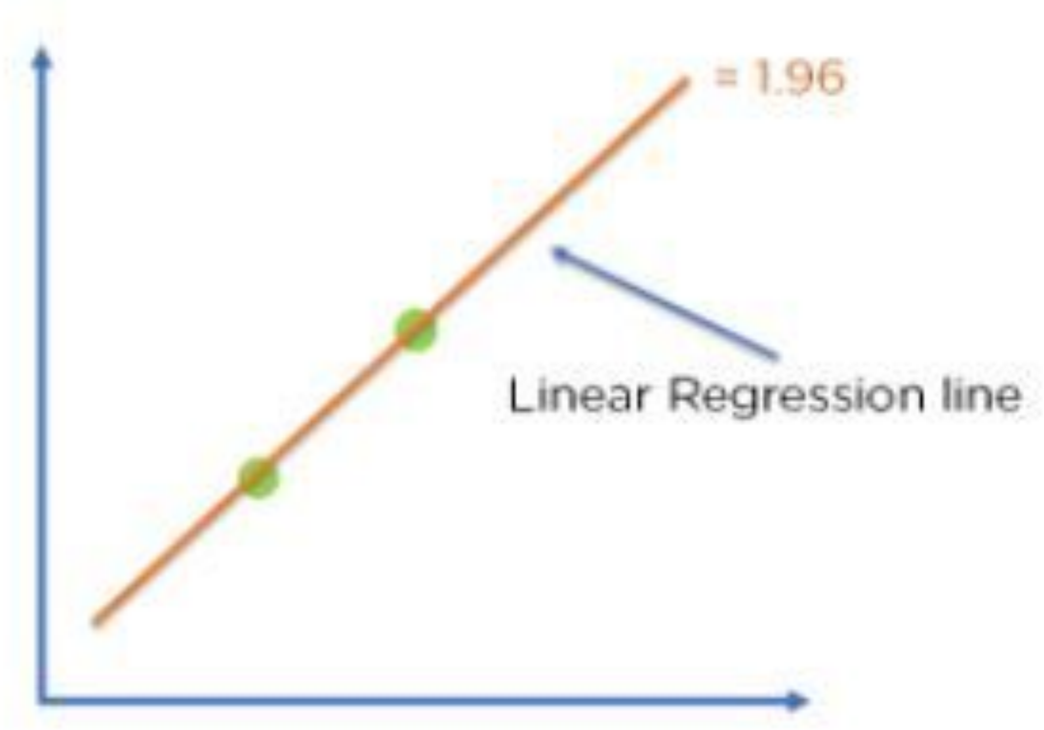
Consider the graph illustrated below which represents Linear regression :

- Cost function = Loss + λ x∑‖w‖^2
- For Linear Regression line, let's consider two points that are on the line,
- Loss = 0 (considering the two points on the line)
- λ= 1
- w = 1.4

Then,

Cost function = 0 + 1 x 1.42

= 1.96



For Ridge Regression, let's assume,

Loss = 0.3^2 + 0.2^2 = 0.13

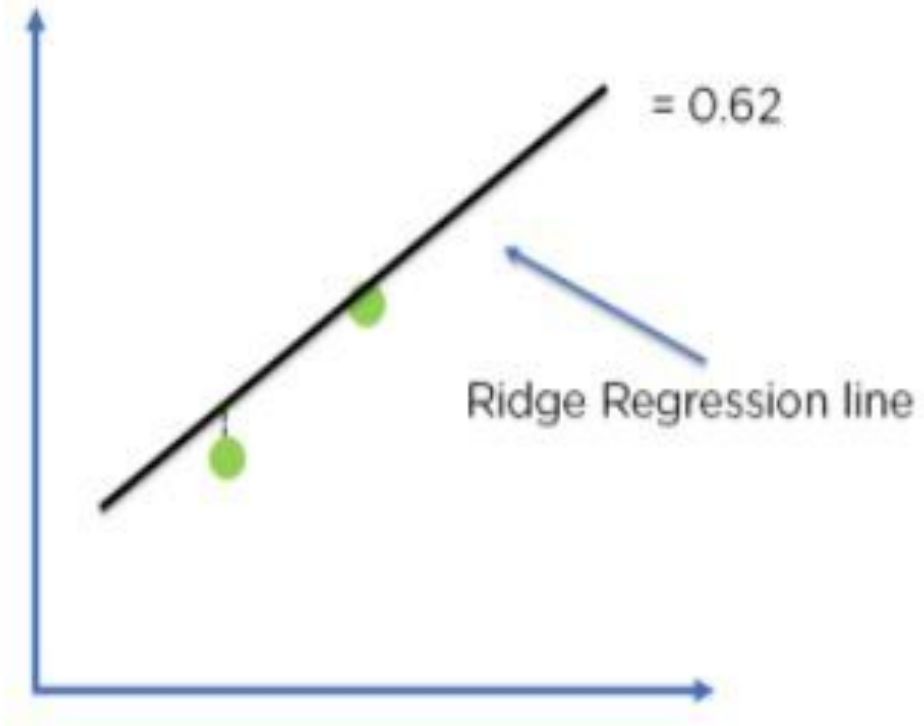
λ = 1

w = 0.7

Then, Cost function = 0.13 + 1 x $(0.7)^2$

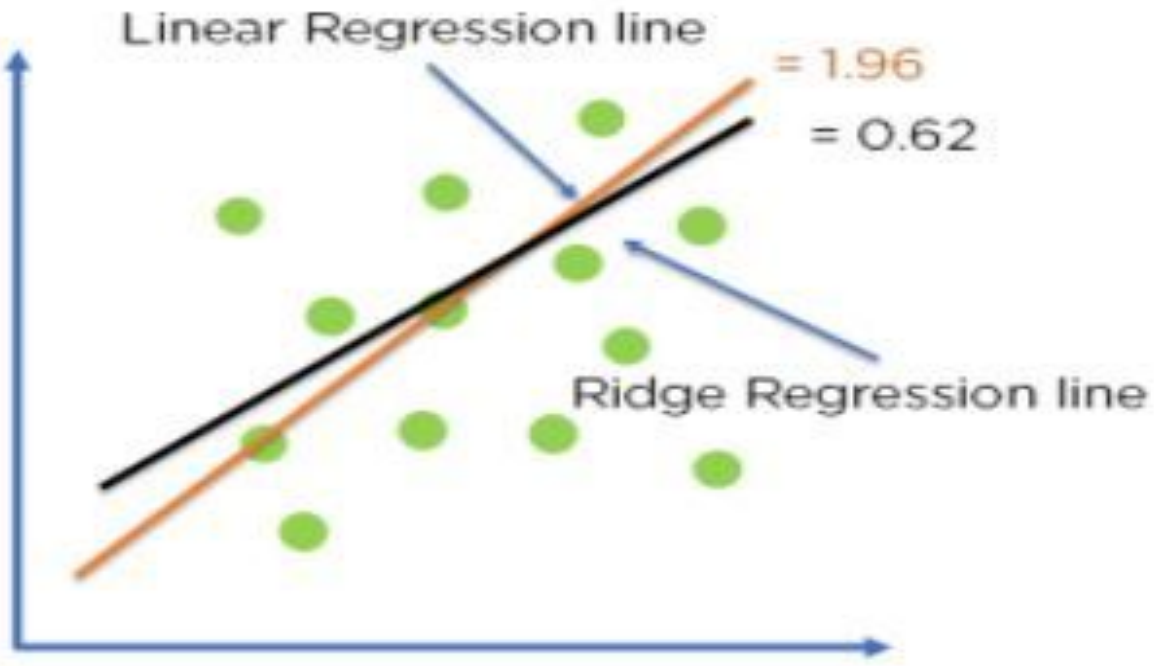
=0.13 + 0.49

= 0.62



Optimization of model fit using Ridge Regression

Comparing the two models, with all data points, we can see that the Ridge regression line fits the model more accurately than the linear regression line.



Lasso Regression

It modifies the over-fitted or under-fitted models by adding the penalty equivalent to the sum of the absolute values of coefficients.

Lasso regression also performs coefficient minimization, but instead of squaring the magnitudes of the coefficients, it takes the true values of coefficients.

Lasso regression, also known as **L1 regularization**, is a linear regression technique that adds a penalty to the loss function to prevent overfitting. This penalty is based on the **absolute values** of the coefficients.

this L1 regularization term reduces overfitting and helps some coefficients to be absolutely zero, hence facilitating feature selection.

The standard loss function (mean squared error) is modified to include a regularization term

Cost function for Lasso Regression

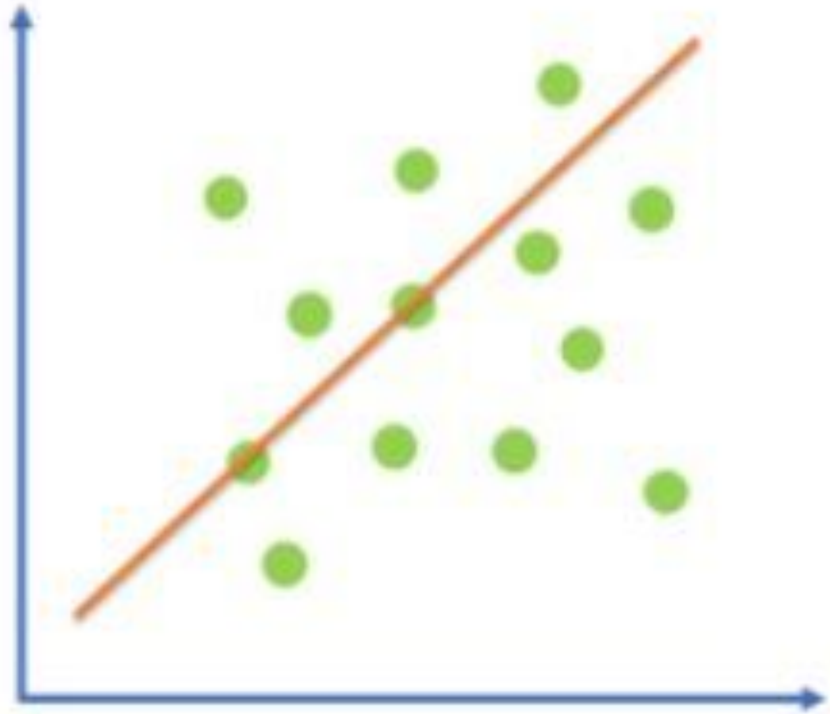
Cost function = Loss + λ x $\sum \|w\|$

Here,

Loss = Sum of the squared residuals

λ = Penalty for the errors

w = slope of the curve/ line



Linear Regression Model

We can control the coefficient values by controlling the penalty terms, just like we did in Ridge Regression. Again consider a Linear Regression model :

Cost function = Loss + λ x ∑‖w‖

For Linear Regression line, let's assume,

Loss = 0 (considering the two points on the line)

λ = 1

w = 1.4

Then, Cost function = 0 + 1 x 1.4

= 1.4

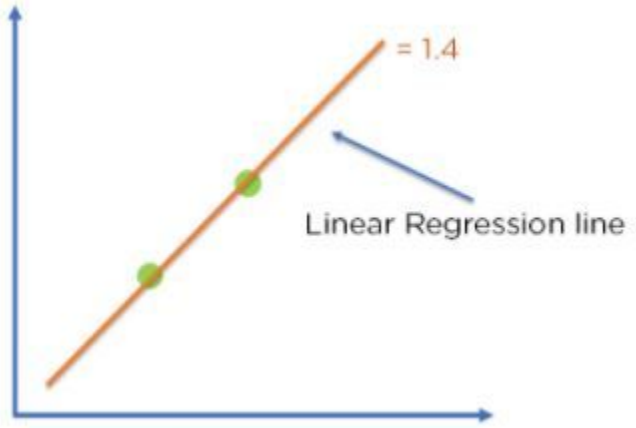


Figure 12: Linear Regression Model

Lasso regression

For Ridge Regression, let's assume

Loss = $0.3^2 + 0.1^2$ = 0.1

λ = 1

w = 0.7

Then, Cost function = 0.1 + 1 x 0.7

= 0.8

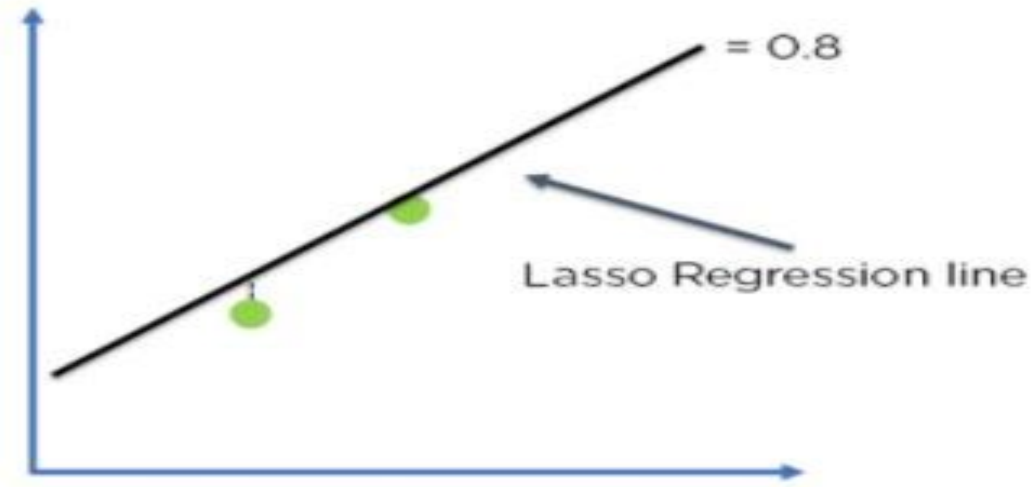


Figure 13: Lasso Regression

Comparing the two models, with all data points, we can see that the Lasso regression line fits the model more accurately than the linear regression line.

What is the difference between Ridge Regression and Linear Regression?

The main difference between ridge regression and linear regression is the addition of a **regularization term** in ridge regression.

In basic linear regression, the model tries to find the best fit by minimizing the error between the predicted and actual values, without any penalty on the size of the coefficients.

What is the difference between Ridge Regression and Lasso Regression?

Ridge regression adds a penalty equal to the square of the coefficient values. This shrinks the coefficients but doesn’t make any of them exactly zero.

While, Lasso regression adds a penalty based on the absolute values of the coefficients. This can shrink some coefficients to zero, effectively removing irrelevant features from the model.

When to Use Ridge Regression?

Ridge Regression is most suitable when all predictors are expected to contribute to the outcome and none should be excluded from the model. It reduces overfitting by shrinking the coefficients, ensuring they don’t become too large, while still keeping all the predictors in the model.

For example, when predicting house prices, features like size, number of bedrooms, location, and year built are all likely relevant. Ridge Regression ensures these features remain in the model but with reduced influence to create a balanced and robust prediction.

When to Use Lasso Regression?

Lasso Regression is ideal when you suspect that only a few predictors are truly important, and the rest may add noise or redundancy. It performs automatic feature selection by shrinking the coefficients of less important predictors to zero, effectively removing them from the model.

For example, in genetic research, where thousands of genes are analyzed for their effect on a disease, Lasso Regression helps by identifying only the most impactful genes and ignoring the irrelevant ones, leading to a simpler and more interpretable model.