🔍 What is Lasso Regression?

Lasso Regression (short for *Least Absolute Shrinkage and Selection Operator*) is a **regularized** version of linear regression.
Regularization helps prevent **overfitting** by adding a penalty to the loss function.

➕ **In Lasso:**

We **add a penalty** based on the **absolute value** of the coefficients (weights), unlike Ridge Regression which uses the square of the coefficients.

➢ Why do we need Lasso?

In linear regression, we try to find weights $w_1, w_2, \ldots, w_n$ such that:

$$\hat{y} = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$$

But if we have **too many features**, the model may:

- Fit noise
- Become complex
- Perform poorly on new data

Lasso helps by:

- **Reducing overfitting**
- **Selecting only important features** (others get zero weight)

🧮 Mathematical Formulation of Lasso

The Lasso loss function is:

$$\text{Loss} = \text{MSE} + \lambda \sum_{j=1}^{n} |w_j|$$

Where:

- $\text{MSE} = \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2$
- λ is a regularization parameter (sometimes called **alpha**)
- $|w_j|$ is the absolute value of the weight

Unlike Ridge, this **L1 norm** can **shrink coefficients to exactly zero**.

🧪 How Are Coefficients Affected by Lambda

- As **lambda increases**, the penalty term grows.
- This **forces more coefficients to become zero**.
- The model becomes **sparser**.

Example:

Lasso(alpha=0.01) → more features retained

Lasso(alpha=1.0) → some weights become 0

Lasso(alpha=100) → most weights go to 0

✅ Great for **feature selection**.

🧠 Higher Coefficients Are Affected More

Lasso affects all coefficients, but:

- **Large weights** contribute more to the regularization term.
- They are more likely to be **shrunk or zeroed** to reduce loss.
- So the model tends to **drop features with large but unstable effects**.

🎯 Impact of Lambda on Bias and Variance

This is related to the **Bias-Variance Tradeoff**.

- **Low λ** → model is flexible → **low bias**, **high variance**
- **High λ** → model is simpler → **high bias**, **low variance**

Lasso with proper λ helps balance both:

- Avoids overfitting
- Keeps the model interpretable

📉 Effect on Loss Function

In Linear Regression:

$$\text{Loss} = \text{MSE}$$

In Lasso Regression:

$$\text{Loss} = \text{MSE} + \lambda \sum |w_j|$$

- The model is penalized **not just for errors** but also for **having too many or too large weights**.
- So Lasso finds a compromise: **small error with fewer or smaller weights**.

➢ Summary of Lasso

Feature	Lasso
Penalty	L1 Norm: (\sum
Effect on Coefficients	Can set them **exactly to 0**
Use Case	Feature selection, sparse models
Bias-Variance	Controls model complexity
Key Parameter	alpha or **λ**