What Is Sparsity in Lasso?

Sparsity means that **many of the model's weights (coefficients) become exactly zero**. This leads to a **simpler model** — one that uses **only the most important features** and ignores the rest.

Why Does Lasso Regression Create Sparsity?

It all comes down to **how Lasso penalizes the weights.**

Lasso Cost Function:

$$J(\theta) = \mathrm{MSE} + \lambda \sum_{j=1}^{n} |\theta_j|$$

- MSE = Mean Squared Error (normal loss)
- λ = regularization strength
- $|\theta_j|$ = absolute value of the coefficients

Compare with Ridge:

- **Ridge** adds $\lambda \sum \theta_j^2$ to the loss — a **squared penalty.**
- **Lasso** adds $\lambda \sum |\theta_j|$ — a **linear/absolute penalty.**

The **shape** of these penalties matters a lot.

Geometry Intuition:

- In **Ridge**, the penalty constraint is **circular (or elliptical)**.
- In **Lasso**, the constraint is **diamond-shaped (like an L1 ball).**

When you minimize the loss with constraints:

- In **Ridge**, the intersection point (minimum) usually falls **inside** the circle → all coefficients small, but **rarely zero**.
- In **Lasso**, the sharp corners of the diamond often touch the loss surface → **many coefficients shrink exactly to 0**.

That's why Lasso creates **sparsity** — its constraint "encourages" zero values.

Practical Interpretation:

- Lasso **automatically performs feature selection**.
- It's very useful when:
    1. You have **many features** (like 100s or 1000s).
    2. Only a **few of them are truly important**.

Summary:

Concept	Ridge	Lasso
Regularization	$\lambda \sum \theta^2$	(\lambda \sum
Coefficients	Shrinks them	Can shrink **some to zero**
Sparsity	No	Yes
Feature Selection	No	Yes
