⚡ Power Transformer | Box-Cox | Yeo-Johnson

These are advanced mathematical methods to transform and normalize data for better model performance.

➢ Why Power Transformations?

Some machine learning models (like linear regression, logistic regression) **perform better when features follow a normal (bell curve) distribution**.
But real-world data is often **skewed** — not evenly spread out.

- That's where **Power Transformers** help:
  They **stabilize variance** and **make the data more normal-like**.

➢ What is PowerTransformer (sklearn)?

sklearn.preprocessing.PowerTransformer is a tool that applies **mathematical transformations** to:

- Remove skewness
- Normalize the data

It supports two main methods:

- **Box-Cox**
- **Yeo-Johnson**

---

1. Box-Cox Transform

- Works **only with positive values**
- Formula:

$$y = \frac{(x^\lambda - 1)}{\lambda} \quad \text{if } \lambda \neq 0$$

- The parameter λ\lambda is chosen automatically to make the data as **normal** as possible

Used when all values are > 0 (e.g., salary, prices, ages)

---

2. Yeo-Johnson Transform

- Works with **both positive and negative values**
- It's a **generalization of Box-Cox**
- Useful when your dataset includes **zeros or negatives**

Ideal for datasets with mixed or negative values

Summary Table:

Method	Handles Negative Values?	Auto-Finds Best λ?	Main Goal
Box-Cox	No	Yes	Normalize & stabilize
Yeo-Johnson	Yes	Yes	Normalize & stabilize