EXPT. NO. NAME

Page No.: Date: YOUVA

M T W T F S S

* Power Transformer:- (Day - 31)

"A Power Transformer is a preprocessing technique that applies power-based mathematical transformations to make data more normally distributed, reduce skewness, and stabilize variance."

⇒ Why it used:-

- To reduce high skewness in data.
- To make distribution more Gaussian-like.
- To improve model performance for algo. sensitive to non-normal data.
- To handle heteroscedasticity (unequal variance)

→ Power Transformer data ne more normal banava mate power-based math function apply kare che.

⇒ Types of Power-Transformers:- [Complex Mathe. Trans.]

(1) Box-Cox Transform:-

"Box-Cox is a power transformation method that transforms strictly positive data into a more normal distribution using an optimized lambda parameter (λ).

Teacher's Signature:

→ Box-Cox only positive data par kaam kare che ane skewness ghatavi ne distribution ne smooth banave che

⇒ Why it is used.

- To reduce positive skewness.
- To stabilize variance.
- To make data more normal for statistical / machine learning models.
- Works well when all values are >0.

⇒ Formula :-

$$x_i^{\lambda} = \begin{cases} \dfrac{x_i^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \ln(x_i) & \text{if } \lambda = 0 \end{cases}$$

⇒ When to use Box-Cox :-

- when all values are strictly positive.
- when your data shows right / positive skewness.
- when you want a transformation similar to log or sqrt but optimized automatically.
- Best for continuous, positive-only features.

→ The exponent here is a variable called lambda (λ) that varies over the range of -5 to 5, and in the process of searching, we examine all values of λ. Finally, we choose the optional value (resulting in the best approximation to a normal distribution) for your variable.

EXPT. NO. | NAME | Page No. | Date | YOUVA

⇒ Aana j special case thi che Log and sq/sqrt.

(2) Yeo-Johnson Transformation :-

"Yeo-Johnson is an extended version of Box-Cox that can transform positive, zero, and negative values without needing to shift the data."

→ Yeo-Johnson positive ane negative banne values par kaam kare che ane data ne normal shape ma lai jai che.

⇒ Why it is used :-

- To reduce skewness in data that has both positive and negative values.
- To automatically estimate the best power parameter (λ).
- To stabilize variance when the dataset contains zeros or negative.

⇒ Formula :-

$$x_i^{(\lambda)} = \begin{cases} [(x_i+1)^{\lambda} - 1]/\lambda & \text{if } \lambda \neq 0,\ x_i \geq 0 \\ \ln(x_i) + 1 & \text{if } \lambda = 0;\ x_i \geq 0 \\ -[(-x_i+1)^{2-\lambda} - 1]/(2-\lambda) & \text{if } \lambda \neq 2,\ x_i < 0 \\ -\ln(-x_i+1) & \text{if } \lambda = 2,\ x_i < 0 \end{cases}$$

Teacher's Signature:

⇒ When to use :-

- When data contains zero or negative values.
- When you want a Box-Cox-like transformation but without positivity restrictions.
- When your data distribution is heavily skewed in either direction.
- Works for both positive and negative continuous feature.

→ This transformation is somewhat of an adjustment to the Box-Cox transformation, by which we can apply it to negative numbers.