Train In Data

Box-Cox transformation

Box-Cox transformation

The Box-Cox transformation is a generalization of the power family of transformations, and it is defined by:

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

where X is the variable and λ is the transformation parameter.

Box-Cox transformation

In the Box-Cox transformation, several values of the parameter λ are evaluated using maximum likelihood, and the λ that returns the best transformation is selected.

The Box-Cox transformation includes all the transformations that we discussed before:

- no transformation (λ = 1),
- the logarithm (λ = 0),
- the reciprocal (λ = -1),
- the square root (when λ = 0.5),
- and the cube root.

Box-Cox transformation

Only suitable for positive variables.

When variables are not strictly positive:

- Add a constant
- Use Yeo-Johnson

Train In Data

THANK YOU

www.trainindata.com