Train In Data

Yeo-Johnson transformation

Yeo-Johnson transformation

The Yeo-Johnson transformation is an extension of the Box-Cox transformation that is no longer constrained to positive values.

In other words, the Yeo-Johnson transformation can be used on variables with zero and negative values as well as positive values.

Yeo-Johnson transformation

The Yeo-Johnson transformation is defined by:

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

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

Yeo-Johnson transformation

- If the variable X is strictly positive, then, the Yeo-Johnson transformation is the same as the Box-Cox power transformation of X + 1.

- If X is strictly negative, then the Yeo-Johnson transformation is the Box-Cox transformation of (-X + 1) but with power 2—λ.

- If the variable has positive and negative values, then the transformation is a mixture of these 2 functions, so different powers are used for the positive and negative values of the variable.

If you ask me, it's a bit of a mess, but as long as it works…

Train In Data

THANK YOU

www.trainindata.com