

Feature scaling

FEATURE SCALING

Scaling methods are those processes acting on ranges of definitions of those features that are used as input by the algorithm. Scaling is a key process in the radiomic workflow not to be under-evaluated since it compromises performances and convergence of the models of ML classifiers if those ranges are not comparable. Among all methods, we find:

- **Linear normalization:** every feature vector X is transformed in a vector X' whose value is comprised between 0 and 1, according to the formula

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2.17)$$

FEATURE SCALING

- **Z-score (Standardization)**: every feature vector X is distributed according to a mean μ and a variance σ is transformed in a vector X' with a standard distribution, with mean 0 and variance 1:

$$X' = \frac{X - \mu}{\sigma}$$

FEATURE SCALING

- **Robust scaler:** every feature vector X is transformed in a vector X' as:

$$X' = \frac{X - X_{median}}{IQR}$$

where IQR is Interquartile Range, the difference between the 1st and the 3rd quartile.

However, only this transformation contributes to strongly reduce the effects due to the outliers, even if it does not eliminate completely.