MACHINE LEARNING

NORMALIZE DATA

Normalization

- Features are usually transformed into a range before the **kNN** algorithm is applied
- **WHY?**
- The distance formula is depend on how features are measured
- If certain features have much larger values than others → the distance measurements will be strongly dominated by the larger values
- We have to rescale the various features such that each one contributes relatively equally to the distance formula
  - **1.)** min-max normalization
  - **2.)** z-transformation

1.) min-max normalization

→ this process transforms a feature such that all of its values fall in a range between **0** and **1**

→ normalized feature values can be interpreted as indicating how far, from **0%** to **100%**, the original value fall along the range between the original minima and maxima

$$X_{new} = \frac{X - \min(X)}{\max(X) - \min(X)}$$

2.) z-score standardization

$$X_{new} = \frac{X - \text{mean}(X)}{StandardDeviation(X)}$$