Overview

- [[Vektor]] as numeric, non-symbolic representation of complex entities
    - complex entity becomes a set of finite numbers
    - no longer express general knowledge
        * hidden within choice of entities and features
- relevant characteristics have to be chosen
    - representable as real numbers
    - ideally easily computable
    - feature/knowledge-engineering
        * know what is actually relevant
        * what questions do we want to ask?
- vector operations
    - identify useful operations in a given use case
    - vector operations can be used to reason over knowledge
- vector representation is easy after feature engineering is done
    - depending on computational complexity
    - can be created automatically
- vector is centered if
    - mean value of vector elements = 0

Useful questions for Feature Engineering

- what are similar entities?
- which group does an entity belong to?
    - classification
- what are meaningful sub-groups?
    - clustering
- is there a correlation between entity characteristics?
    - [[Correlation]]
- does a characteristic cause another one?
    - causation
- is there a more compact representation?
    - factor analysis
    - which variables carry most information?

Similarity

- useful for
    - [[Recommender Systems]]

- – [[Information Retrieval]]
- – [[Classification]]
- – clustering
- many types of simillarity measures
- cosine similarity
  - – similarity = angle between two vectors
    - * length independent
      - ◆ normalized vectors
    - * direction dependent
    - * $cos : [0, 360] --> [-1, 1]$
    - * $sim(a, b) = cos(\phi) = \frac{a*b}{||a||*||b||}$
  - – angles
    - * 0° ==> extremely similar
      - ◆ no equality
    - * 90° ==> orthogonal, not similiar at all
    - * 180° ==> opposite direction, inverse

[[Knowledge Representation]]