

Fundamentals of Machine Learning

Distances and Similarity

Jan Platos

November 25, 2021

Distances and Similarity

Distances and Similarity

- How far apart are the objects?
- How close together are the objects?

Distances and Similarity - Distance

- Distance usually refers to a category of functions that measures difference in a Cartesian space.
- The smaller distances the more closer the objects are.
- To maintain basic properties of the coordinate system, we expect the function to be a Metric:
 1. $d(x, y) \geq 0$ (non-negativity)
 2. $d(x, y) = 0 \Leftrightarrow x = y$ (identity)
 3. $d(x, y) = d(y, x)$ (symmetry)
 4. $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality)

Distances and Similarity - Distance

- Manhattan distance (Taxi Driver/City Block)

$$d(X, Y) = \sum_{i=1}^N |x_i - y_i|$$

- Euclidean distance

$$d(X, Y) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2}$$

Distances and Similarity - Similarity

- Similarity measures the closeness between objects.
- The higher values the more closer the objects are.
- Similarity does not require Cartesian coordinates.
- Object-related similarity measures may be defined.

Distances and Similarity - Similarity

- Cosine Similarity measures the angle between objects in a euclidean space.

$$\cos(X, Y) = \frac{X \cdot Y}{\|X\| \|Y\|}$$

Questions