

Knowledge Discovery in Databases
 WS 2010/11

Übungsblatt 2: Distanzmaße

Aufgabe 2-1 Distanzmaße

Distanzmaße können wir nach ihren Eigenschaften in folgende Kategorien einteilen:

$d : S \times S \rightarrow \mathbb{R}_0^+$ $x, y, z \in S :$	reflexiv $x = y \Rightarrow d(x, y) = 0$	symmetrisch $d(x, y) = d(y, x)$	strikt $d(x, y) = 0 \Rightarrow x = y$	Dreiecks-Ungleichung $d(x, z) \leq d(x, y) + d(y, z)$
Unähnlichkeitsfunktion	×			
Distanzfunktion	×	×		
Semi-Metrik	×	×	×	
Pseudo-Metrik	×	×		×
Metrik	×	×	×	×

D.h., wenn ein Distanzmaß $d : S \times S \rightarrow \mathbb{R}_0^+$ für alle $x, y, z \in S$: reflexiv, symmetrisch und strikt ist sowie die Dreiecks-Ungleichung erfüllt, ist es eine Metrik. Wie Sie sehen, muß eine Distanzfunktion nicht *strikt* reflexiv sein. Machen Sie sich den Unterschied zwischen Reflexivität und Striktheit klar!

Anmerkung: Die Namen Distanzfunktion, Semi- und Pseudo-Metrik werden in der Literatur nicht einheitlich definiert.

Entscheiden Sie für die folgenden Funktionen $d(\mathbb{R}^n, \mathbb{R}^n)$ jeweils, ob es sich um ein Distanzmaß handelt, und wenn ja, in welche Kategorie es fällt.

(a) $d(x, y) = \sum_{i=1}^n (x_i - y_i)$

(b) $d(x, y) = \sum_{i=1}^n (x_i - y_i)^2$

(c) $d(x, y) = \sqrt{\sum_{i=1}^{n-1} (x_i - y_i)^2}$

(d) $d(x, y) = \sum_{i=1}^n \begin{cases} 1 & \text{falls } x_i = y_i \\ 0 & \text{falls } x_i \neq y_i \end{cases}$

(e) $d(x, y) = \sum_{i=1}^n \begin{cases} 1 & \text{falls } x_i \neq y_i \\ 0 & \text{falls } x_i = y_i \end{cases}$