

Prof. Dr. A. Voisard, N. Lehmann

Datenbanksysteme, SoSo 20

Übung 07

TutorIn: Gröling, Marc

Tutorium 04

David Ly & Thore Brehmer

10. Juni 2020

2 Aufgabe: Schlüssel

(30 Punkte)

Datenpunkte	A	B	C	D	E	F	Clusterzentren	G	H
X	1	1	2	-3	-2	0	X	1	-3
Y	3	2	1	1	-1	-2	Y	1	-2

Tabelle 2: Datenpunkte und Clusterzentren

- 1) Führen Sie mit den in Tabelle 2 gegebenen und in Abbildung 1 visualisierten Datenpunkten und Clusterzentren den *k*-means Algorithmus aus. Berechnen Sie, welche Punkte zu welchen Clusterzentren gehören. Verwenden Sie als Distanzmaß die Manhattan-Distanz. Der Algorithmus terminiert, wenn sich die Clusterzentren weniger als $\delta = 0,25$ bewegen. (25 P.)

1. Iteration:

Distanz von G zu: A=2, B=1, C=1, D=4, E=5, F=4

Distanz von H zu: A=9, B=8, C=8, D=3, E=2, F=3

Zu Clusterzentrum G gehören die Punkte: {A, B, C}

Zu Clusterzentrum H gehören die Punkte: {D, E, F}

Neue Clusterzentren sind:

Clusterzentren	G	H
X	1.3	-1.6
Y	2	-0.6
moved by	1.3	2.6

2. Iteration:

Distanz von G zu: $A=1.\bar{3}$, $B=0.\bar{3}$, $C=1.\bar{6}$, $D=5.\bar{3}$, $E=6.\bar{3}$, $F=5.\bar{3}$

Distanz von H zu: $A=6.\bar{3}$, $B=5.\bar{3}$, $C=5.\bar{3}$, $D=3$, $E=0.\bar{6}$, $F=3$

Zu Clusterzentrum G gehören die Punkte: {A, B, C}

Zu Clusterzentrum H gehören die Punkte: {D, E, F}

Neue Clusterzentren sind:

Clusterzentren	G	H
X	$1.\bar{3}$	$-1.\bar{6}$
Y	2	$-0.\bar{6}$
moved by	0	0

Terminierung

2) *Wie viele Iterationen durchläuft der Algorithmus?*

(5 P.)

Zwei Iterationen, danach terminiert der Algorithmus.

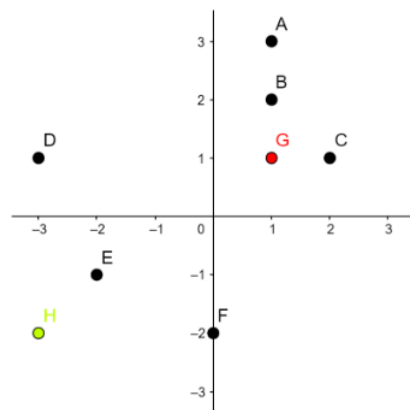


Abbildung 1: Visualisierung der Datenpunkte und Clusterzentren