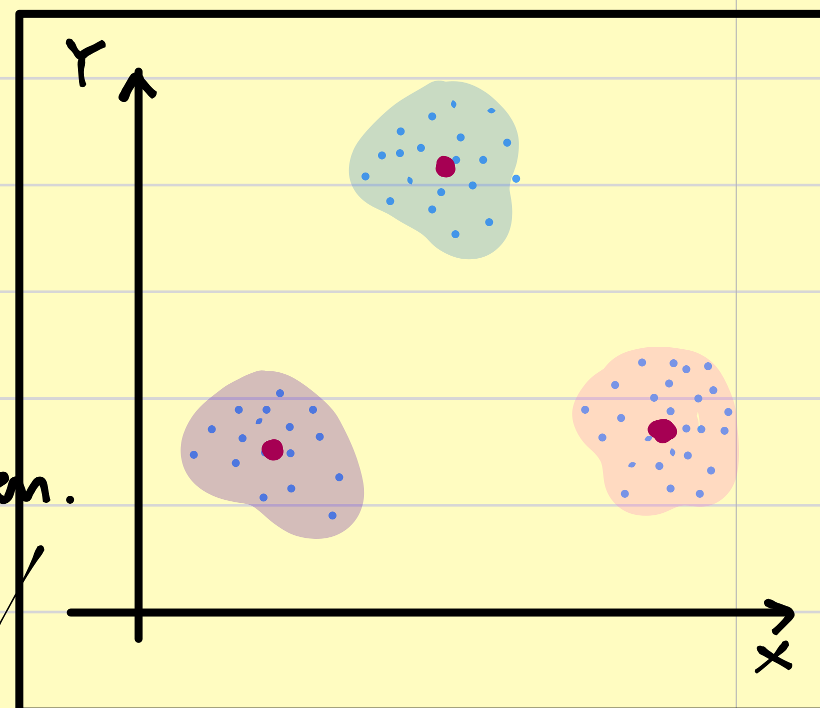
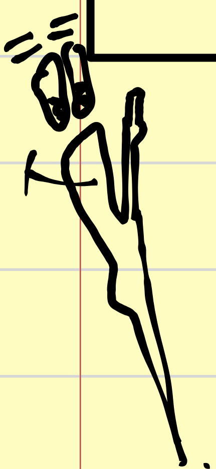
20251205_Mathematik1_B

MASCHINE LEARNING

K-MEANS CLUSTERING

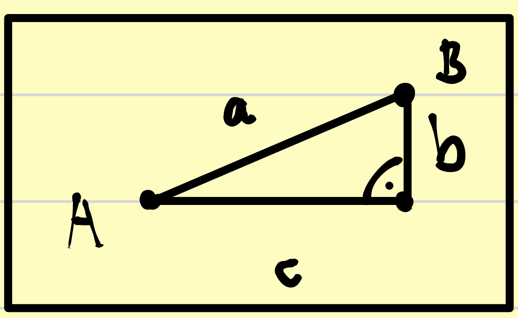


Hypothese: Ähnliche Datensätze sind in der „Nähe" voneinander.

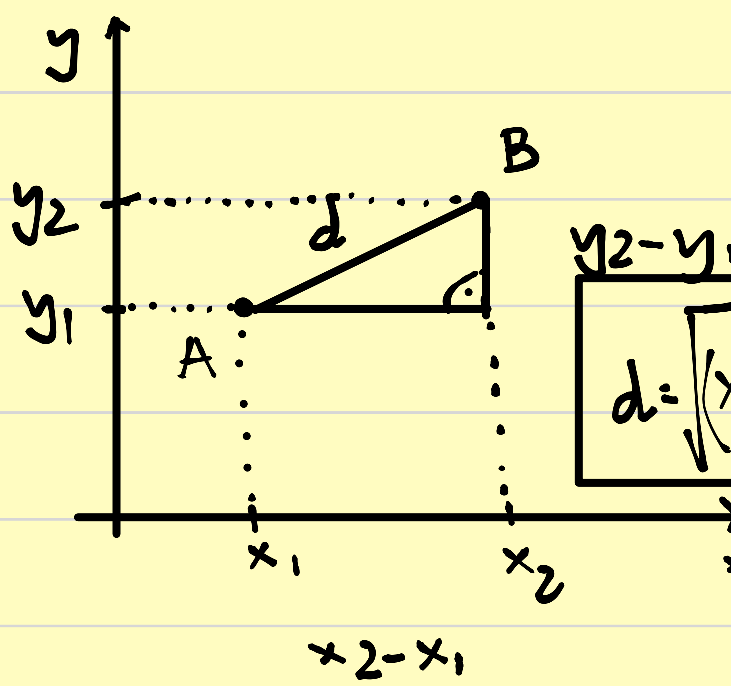


- wir können einen Abstand messen.
- wir wissen wie viele CLUSTERS / GRUPPEN es gibt (K).

Abstand zw zwei Punkten wird mittels $a^2 = b^2 + c^2$ geliefert



$a^2 = b^2 + c^2 \rightarrow a = \sqrt{b^2 + c^2}$



$y_2 - y_1$

$$\boxed{d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}$$

„K" bedeutet die Anzahl Gruppen (MUSS VOM NUTZER VORGEGEBEN WERDEN)

Schritte: #0. Entscheidung über Anzahl Clusters „K"

#1. Punkte werden in „K" Gruppen geteilt.

#2. Zentroide (Schwerpunkte) der Gruppen werden ermittelt.

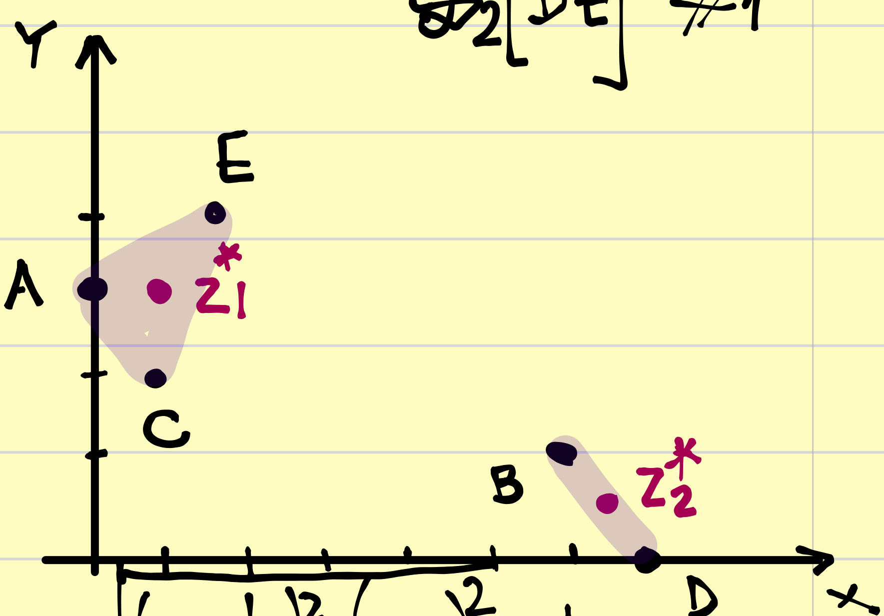
#3. Abstand von den Zentroiden zu den Punkten wird berechnet.

#4. Gruppenbildung nach dem geringsten Abstand.
#5. Zurück zu #1 bis die Gruppen konstant bleiben.

Beispiel. Gegeben sind die [X,Y] Koordinaten von 5 Werken. Bitte ermitteln Sie die optimale Position von 2 Lager, angenommen Alle Werke haben den gleichen Bedarf.

	A	B	C	D	E
X:	0	6	1	7	2
Y:	3	1	2	0	4

Anfangsgruppen: G_1[ABC] #0
G_2[DE] #1



#2. $Z_1: \left[\frac{0+6+1}{3}, \frac{3+1+2}{3}\right] = [2'33, 2]$

$Z_2: \left[\frac{7+2}{2}, \frac{0+4}{2}\right] = [4'5, 2]$

#3. $d_{AZ1} = \sqrt{(0-2'33)^2+(3-2)^2} = 2'53 < d_{AZ2} = \sqrt{(0-4'5)^2+(3-2)^2} = 4'6$

$d_{BZ1} = \sqrt{(6-2'33)^2+(1-2)^2} = 3'8 > d_{BZ2} = \sqrt{(6-4'5)^2+(1-2)^2} = 1'8$

$d_{CZ1} = \sqrt{(1-2'33)^2+(2-2)^2} = 1'33 < d_{CZ2} = \sqrt{(1-4'5)^2+(2-2)^2} = 3'5$

$d_{DZ1} = \sqrt{(7-2'33)^2+(0-2)^2} = 5'08 > d_{DZ2} = \sqrt{(7-4'5)^2+(0-2)^2} = 3'2$

$d_{EZ1} = \sqrt{(2-2'33)^2+(4-2)^2} = 2'02 < d_{EZ2} = \sqrt{(2-4'5)^2+(0-2)^2} = 3'2$

Gruppen: G_1^* [ACE] G_2^* [BD] #1

#2. $Z_1^*: \left[\frac{0+1+2}{3}, \frac{3+2+4}{3}\right] = [1,3]$ $Z_2^*: \left[\frac{6+7}{2}, \frac{1+0}{2}\right] = [6'5, 0'5]$

#3. $d_{AZ_1^*} = \sqrt{(0-1)^2+(3-3)^2} = 1 < d_{AZ_2^*} = \sqrt{(0-6'5)^2+(3-0'5)^2} = 6'9$

$d_{CZ_1^*} = \sqrt{(1-1)^2+(2-3)^2} = 1 < d_{CZ_2^*} = \sqrt{(1-6'5)^2+(2-0'5)^2} = 5'7$

$d_{EZ_1^*} = \sqrt{(2-1)^2+(4-3)^2} = 0'707 < d_{EZ_2^*} = \sqrt{(2-6'5)^2+(4-0'5)^2} = 5'7$

$d_{BZ_1^*} = \sqrt{(6-1)^2+(1-3)^2} = 5'38 > d_{BZ_2^*} = \sqrt{(6-6'5)^2+(1-0'5)^2} = 0'707$

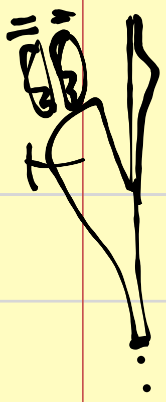
$d_{DZ_1^*} = \sqrt{(7-1)^2+(0-3)^2} = 6'7 > d_{DZ_2^*} = \sqrt{(7-6'5)^2+(0-0'5)^2} = 0'707$

Lösung: Läger $Z_1^*[1,3]$ $Z_2^*[6'5, 0'5]$
Gruppen [ACE] [B,D]

Beispiel. Die Positionen von 6 Werken mit unterschiedlichen Bedarfen an Rohwahre sind durch Ihre Koordinaten auf der Karte bestimmt.
Jedes Werk wird von einem der 2 geplanten Lägern beliefert. Um die Fahrtkosten zu minimieren sollten die Läger so positioniert werden, dass sowohl die Werke möglichst nah sind, als auch die Bedarfe berücksichtigt werden. Bitte sprechen Sie der GF eine Empfehlung aus, für die Lagerpositionen.

	A	B	C	D	E	F
X	1	2	0	6	7	3
Y	3	2	1	1	2	3
B	2	1	3	1	3	1

Anfangsgruppen
$G_1[ACD]$ $G_2[BEF]$



HINWEIS: Gewichteter Mittelwert

$$z_1 = \left[\frac{1\cdot 2 + 0\cdot 3 + 6\cdot 1}{2+3+1}, \frac{3\cdot 2 + 1\cdot 3 + 1\cdot 1}{2+3+1}\right] = [\ldots, \ldots]$$