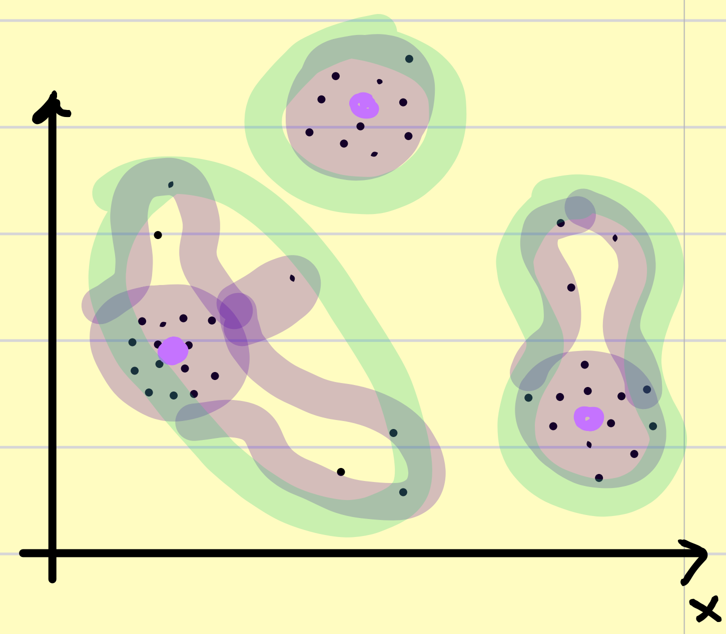
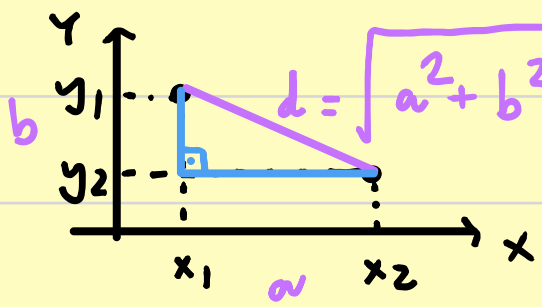
MASCHINELLES LERNEN

K-Means Clustering



- „Clustering" bedeuten Gruppenbildung.
- „Ähnliche Punkte" sind Punkt die in der Nähe voneinander liegen.



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

(Pythagoras)

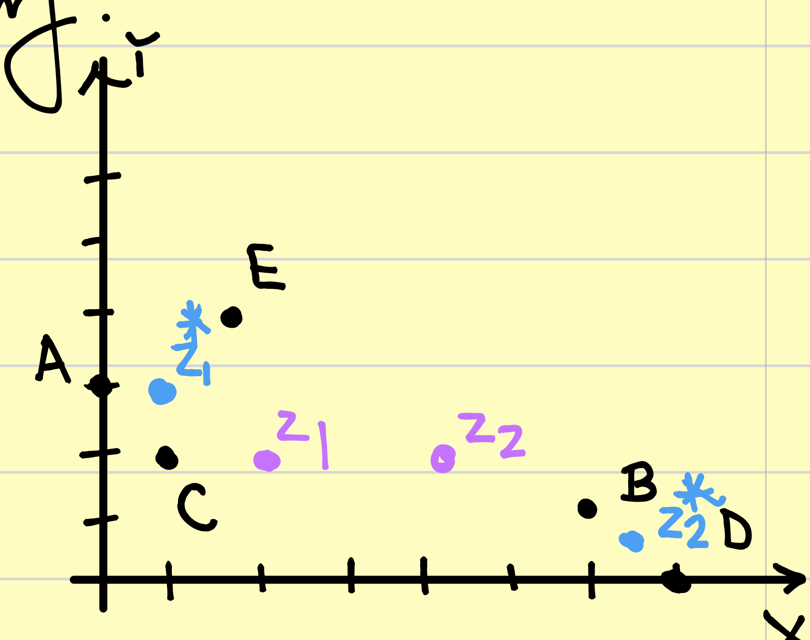
- Die Hypothese um K-Means Clustering anzuwenden ist, dass der von den Daten definierten Raum eine Euklydische Natur hat: wir können einen Abstand messen.
- „K" sind die Anzahl Gruppen und müssen vom Nutzer vorgegeben werden.
- K-Means Clustering zeigt uns die Position der Punkte mit den geringsten Abstand zu den Zentren (•) = Zentroide.
- NACHTEIL: wir müssen vorab entscheiden, wie viele Zentren.
- VORTEIL: schnell & effizient.

k-Means Clustering Schritte:

- Schritt 0. Entscheidung über Anzahl Clusters (K).
- Schritt 1. Punkte von Dataset werden in K Gruppen geteilt.
- Schritt 2. Zentroide (Schwerpunkt) der Gruppen werden ermittelt.
- Schritt 3. Abstand von den Punkten zu den Zentroiden wird berechnet.
- Schritt 4. Gruppenbildung (Clustern) nach dem geringsten Abstand und neu bei Schritt 1. anfangen bis Abstand zu ~~den~~ Zentroiden konstant ist.

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.

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



- Bitte fangen Sie mit den Gruppen G_1 [A,B,C] G_2 [D,E]

Schritt 0. Sind die Daten Euklydisch? ✓
Clusters k=2.

Schritt 1. $G_1[A,B,C]$ $G_2[D,E]$ ✓

Schritt 2. $z_1 = \left[\frac{0+6+1}{3}, \frac{3+1+2}{3}\right] = [2'33, 2]$; $z_2 = \left[\frac{7+2}{2}, \frac{0+4}{2}\right] = [4'5, 2]$

Schritt 3.

$d_{A,Z_1} = \sqrt{(0-2'33)^2+(3-2)^2} = 2'535$ $\qquad$ $d_{A,Z_2} = \sqrt{(0-4'5)^2+(3-2)^2} = 4'609$

$d_{B,Z_1} = \sqrt{(6-2'33)^2+(1-2)^2} = 3'804$ $\qquad$ $d_{B,Z_2} = \sqrt{(6-4'5)^2+(1-2)^2} = 1'803$

$d_{C,Z_1} = \sqrt{(1-2'33)^2+(2-2)^2} = 1'33$ $\qquad$ $d_{C,Z_2} = \sqrt{(1-4'5)^2+(2-2)^2} = 3'5$

$d_{D,Z_1} = \sqrt{(7-2'33)^2+(0-2)^2} = 5'08$ $\qquad$ $d_{D,Z_2} = \sqrt{(7-4'5)^2+(0-2)^2} = 3'2$

$d_{E,Z_1} = \sqrt{(2-2'33)^2+(4-2)^2} = 2'027$ $\qquad$ $d_{E,Z_2} = \sqrt{(2-4'5)^2+(4-2)^2} = 3'2$

Schritt 4. Neue Gruppenbildung.

$G_1^*[A,C,E]$ $\qquad$ $G_2^*[B,D]$

$\shortparallel$

Schritt 1^*.

Schritt 2^*. $z_1^* = \left[\frac{0+1+2}{3}, \frac{3+2+4}{3}\right] = [1,3]$ $\qquad$ $z_2^* = \left[\frac{6+7}{2}, \frac{1+0}{2}\right] = [6'5, 0'5]$

Schritt 3^*.

$d_{A,z_1^*} = \sqrt{(0-1)^2+(3-3)^2} = 1$ $\qquad$ $d_{A,z_2^*} = \sqrt{(0-6'5)^2+(3-0'5)^2} = 6'9$

$d_{C,z_1^*} = \sqrt{(1-1)^2+(2-3)^2} = 1$ $\qquad$ $d_{C,z_2^*} = \sqrt{(1-6'5)^2+(2-0'5)^2} = 5'7$

$d_{E,z_1^*} = \sqrt{(2-1)^2+(4-3)^2} = 0'707$ $\qquad$ $d_{E,z_2^*} = \sqrt{(2-6'5)^2+(4-0'5)^2} = 5'7$

$d_{B,z_1^*} = \sqrt{(6-1)^2+(1-3)^2} = 5'38$ $\qquad$ $d_{B,z_2^*} = \sqrt{(6-6'5)^2+(1-0'5)^2} = 0'707$

$d_{D,z_1^*} = \sqrt{(7-1)^2+(0-3)^2} = 6'7$ $\qquad$ $d_{D,z_2^*} = 0'707$

Position der Läger $z_1^*[1,3]$ $z_2^*[6'5, 0'5]$
Gruppen: [A, C, E] [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 nutzen Sie einen geeigneten Algorythmus um der Geschäftsführung eine Empfehlung für die Lagepositionen auszusprechen.

Daten: X [1, 2, 0, 6, 7, 3]
Y [3, 2, 1, 1, 2, 3]
Bedarfe [2, 1, 3, 1, 3, 1]