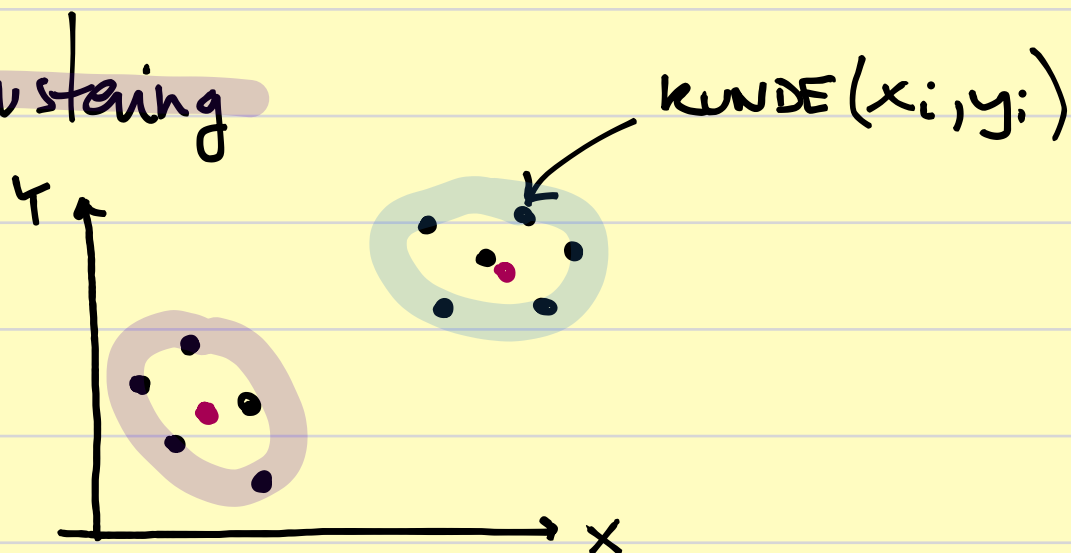


k-Means Clustering

Beispiel:



- k sind die Anzahl Gruppen und k-Means Clustering zeigt uns die Position der Punkte mit den geringsten Abstand zu den Gruppen
- NACHTEIL: wir müssen dem Algorithmus sagen, wie viele Gruppen wir haben
- VORTEIL: schnell und effizient.

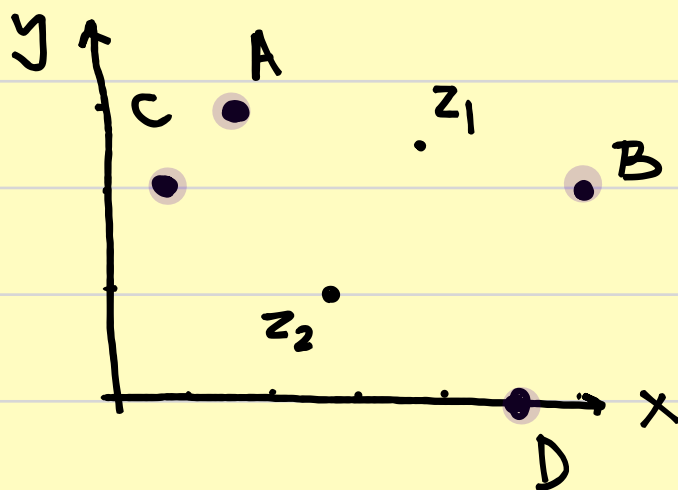
K-MEANS CLUSTER ALGORITHMUS

SCHRITT 0. ENTSCHEIDEN #CLUSTERS k .

- SCHRITT 1. PUNKTE VOM DATASET IN k GRUPPENTEILEN.
- SCHRITT 2. ZENTROID (SCHWERPUNKT) DER GRUPPEN ERMITTELN
- SCHRITT 3. ABSTAND VON DEN PUNKTEN ZUM ZENTROID.
- SCHRITT 4. CLUSTERN NACH GERINGSTEN ABSTAND UND NEU BEI SCHRITT 1. ANFANGEN.
- SCHRITT 5. bis ABSTAND ZUDEN ZENTROIDEN KONSTANT IST.

Beispiel.

	x	y
A	2	3
B	6	2
C	1	2
D	5	0



SCHRITT 0. WIR SUCHEN 2 CLUSTERS.

SCHRITT 1. PUNKTE VOM DATASET IN 2 GRUPPEN TEILEN

$$\{A, B\} \quad \{C, D\}$$

SCHRITT 2. ZENTROIDE .

$$1. \{A, B\} \quad \left. \begin{array}{l} x_1 = \frac{2+6}{2} = 4 \\ y_1 = \frac{3+2}{2} = 2.5 \end{array} \right\} z_1 = (4, 2.5)$$

$$2. \{C, D\} \quad \left. \begin{array}{l} x_2 = \frac{1+5}{2} = 3 \\ y_2 = \frac{2+0}{2} = 1 \end{array} \right\} z_2 = (3, 1)$$

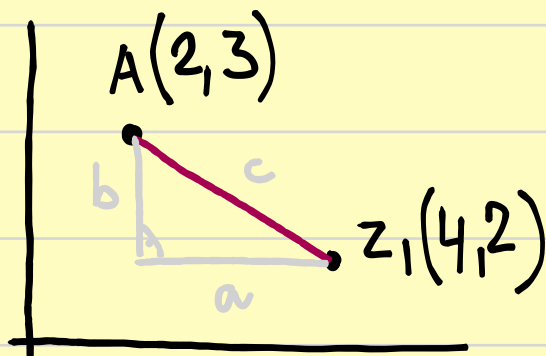
SCHRITT 3. ABSTAND PUNKTE ZUM ZENTROID

$$d(A, z_1) = \sqrt{(2-4)^2 + (3-2.5)^2} = 2.06$$

$$d(B, z_1) = \sqrt{(6-4)^2 + (1-2.5)^2} = 2.5$$

$$d(C, z_1) = \sqrt{(1-4)^2 + (2-2.5)^2} = 3.04$$

$$d(D, z_1) = \sqrt{(5-4)^2 + (0-2.5)^2} = 2.69$$



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

$$d(A, z_2) = \sqrt{(2-3)^2 + (3-1)^2} = 2.23$$

$$d(B, z_2) = \sqrt{(6-3)^2 + (2-1)^2} = 3.16$$

$$d(C, z_2) = \sqrt{(1-3)^2 + (2-1)^2} = 2.23$$

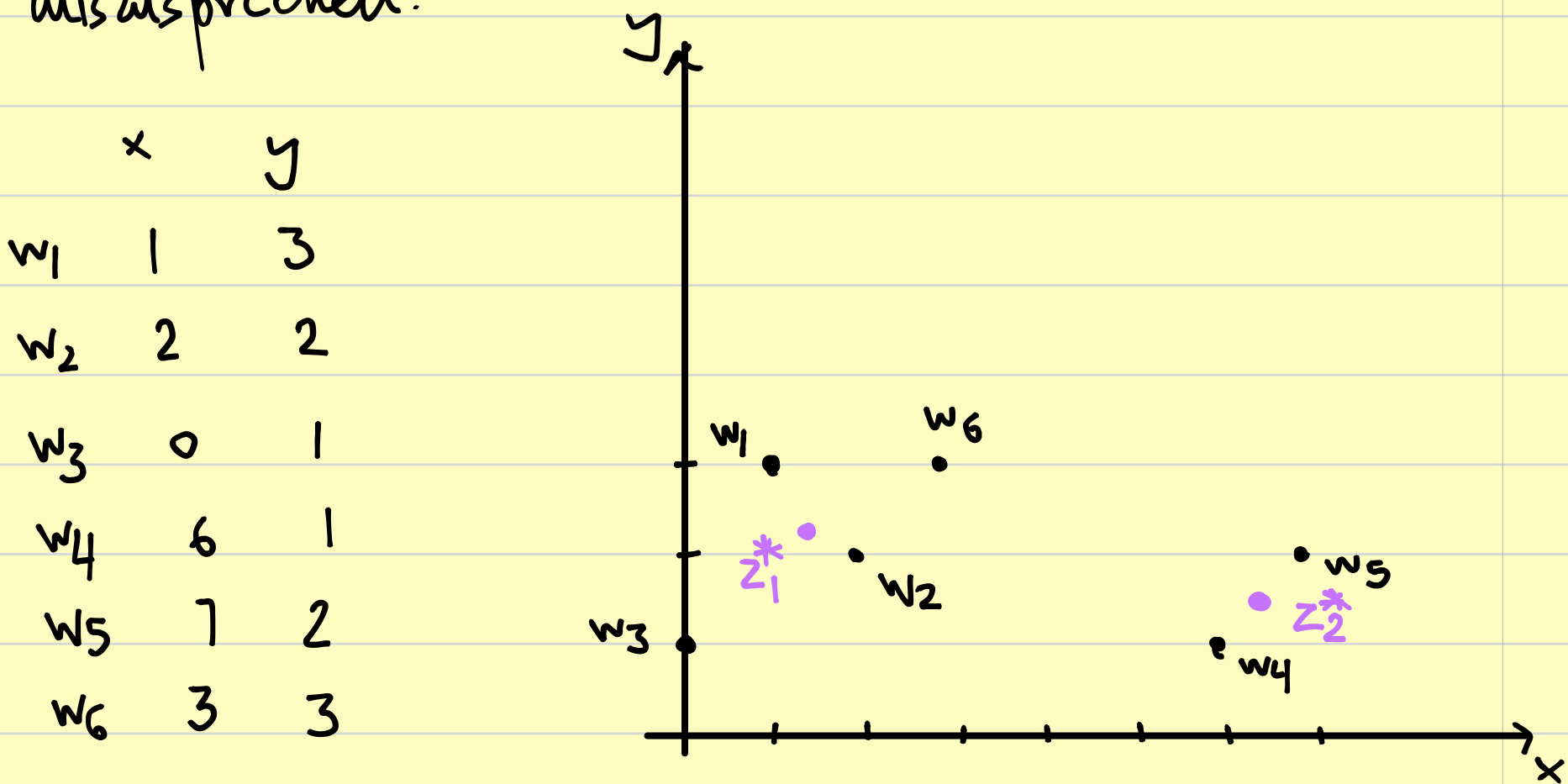
$$d(D, z_2) = \sqrt{(5-3)^2 + (0-1)^2} = 2.23$$

1. (A, B)

2. (C, D)

Beispiel. Die Positionen von 6 WERKEN sind durch ihre Koordinaten auf der Karte bestimmt. Jedes Werk wird von einem der zwei geplanten Lagern beliefert. Um die Fahrtkosten zu minimieren sollten die Lager so positioniert werden, dass die Werke möglichst nah sind.

Bitte nutzen sie einen geeigneten Algorithmus um der Geschäftsführung eine Empfehlung für die Lagerpositionierung auszusprechen.



SCHRITT 0. 2 GRUPPEN

SCHRITT 1. 1. {w₁, w₂, w₃} 2. {w₄, w₅, w₆}

SCHRITT 2.

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

$$z_2 = \left[\frac{6+7+3}{3}, \frac{1+2+3}{3} \right] = [5.33, 2]$$

SCHRITT 3.

$$d(w_1, z_1) = \sqrt{(1-1)^2 + (3-2)^2} = 1$$

$$d(w_2, z_1) = \sqrt{(2-1)^2 + (2-2)^2} = 1$$

$$d(w_3, z_1) = \sqrt{(0-1)^2 + (1-2)^2} = 1$$

$$d(w_4, z_1) = \sqrt{(6-1)^2 + (1-2)^2} = 5'10$$

$$d(w_5, z_1) = \sqrt{(7-1)^2 + (2-2)^2} = 6$$

$$d(w_6, z_1) = \sqrt{(3-1)^2 + (3-2)^2} = 2'24$$

$$d(w_1, z_2) = \sqrt{(1-5'33)^2 + (3-2)^2} = 4'44$$

$$d(w_2, z_2) = \sqrt{(2-5'33)^2 + (2-2)^2} = 3'33$$

$$d(w_3, z_2) = \sqrt{(0-5'33)^2 + (1-2)^2} = 5'42$$

$$d(w_4, z_2) = \sqrt{(6-5'33)^2 + (1-2)^2} = 1'20$$

$$d(w_5, z_2) = \sqrt{(7-5'33)^2 + (2-2)^2} = 1'67$$

$$d(w_6, z_2) = \sqrt{(3-5'33)^2 + (3-2)^2} = 2'54$$

SCHRITT 4.

NEUE GRUPPEN: 1. $\{w_1, w_2, w_3, w_6\}$ 2. $\{w_4, w_5\}$

Da sich die Gruppen geändert haben, müssen wir die Schritte wiederholen.

SCHRITT 2.

$$z_1^* = \left[\frac{1+2+0+3}{4}, \frac{3+2+1+3}{4} \right] = [1'5, 2'25]; \quad z_2^* = \left[\frac{6+7}{2}, \frac{1+2}{2} \right] = [6'5, 1'5]$$

SCHRITT 3.

$$d(w_1, z_1^*) = \sqrt{(1-1'5)^2 + (3-2'25)^2} < d(w_1, z_2^*) = \sqrt{(1-6'5)^2 + (3-1'5)^2}$$

$$d(w_2, z_1^*) = \sqrt{(2-1'5)^2 + (2-2'25)^2} < d(w_2, z_2^*) = \sqrt{(2-6'5)^2 + (2-1'5)^2}$$

$$d(w_3, z_1^*) < d(w_3, z_2^*)$$

$$d(w_4, z_1^*) > d(w_4, z_2^*)$$

$$d(w_5, z_1^*) > d(w_5, z_2^*)$$

$$d(w_6, z_1^*) < d(w_6, z_2^*)$$

GRUPPEN bleiben und das algorithmus ist beendet. ✓
(CLUSTERS)

ANTWORT: 1. $\{w_1, w_2, w_3, w_6\}$ 2. $\{w_4, w_5\}$

$$\text{Lager: } z_1^* = [1'5, 2'25] \quad z_2^* = [6'5, 1'5]$$

Übung. Die Werte der oberen Übung haben unterschiedliche Bedarfe. Bitte kalkulieren Sie die Position der Lager erneut um dies zu berücksichtigen.

Bedarfe $\{ 2, 1, 2, 1, 2, 3 \}$
 $w_1 \quad w_2 \quad w_3 \quad w_4 \quad w_5 \quad w_6$

Hinweis: arbeiten Sie mit gewichteter Zentralide.

