

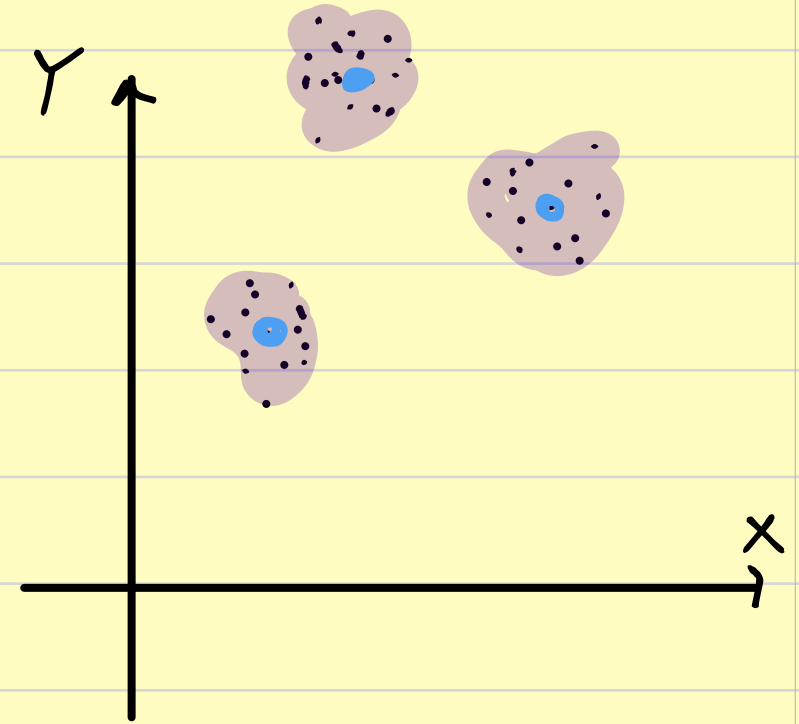
Theorie

kw19 [0321][0328] Theorie
kw20 [0404][0411] Theorie
kw21 [0416][0423] Theorie
kw22 [0502][0514] Theorie

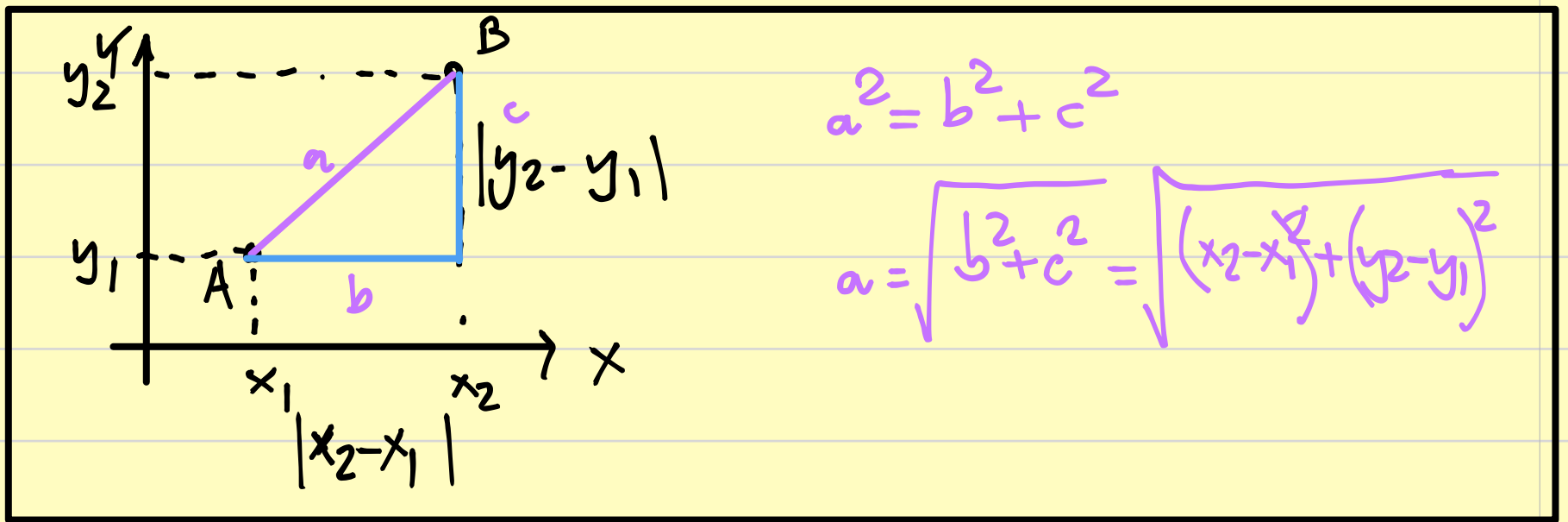
kw23 [0521][0528] Übungen
kw24
kw25 Probeklausur Übungen
kw26
kw27 Prüfung

Maschinelles Lernen . k-Means clustering

Wofür ist die nützlich?
• Eine Spedition möchte seinen Kunden aus 3 städte mit 3 Lieferwagen versorgen. Die räumliche position der Kunden ist bekannt. Wo sollten die lade-Stationen platziert werden für die fahrzeuge?



- ..Clustering bedeutet Gruppenbildung.
- Gruppen werden anhand der Ähnlichkeit der Daten gebildet.
- Ähnlich bedeutet in der Nähe im Datenraum.



- .."k" sind die Anzahl Gruppen und müssen vorgegeben werden.

K-Means Clustering



• Schritt 0. Entscheidung über die Anzahl clusters (k)

→ • Schritt 1. Punkte vom Dataset in den k Gruppen verteilen.

• Schritt 2. Schwerpunkte der Gruppen ermitteln.

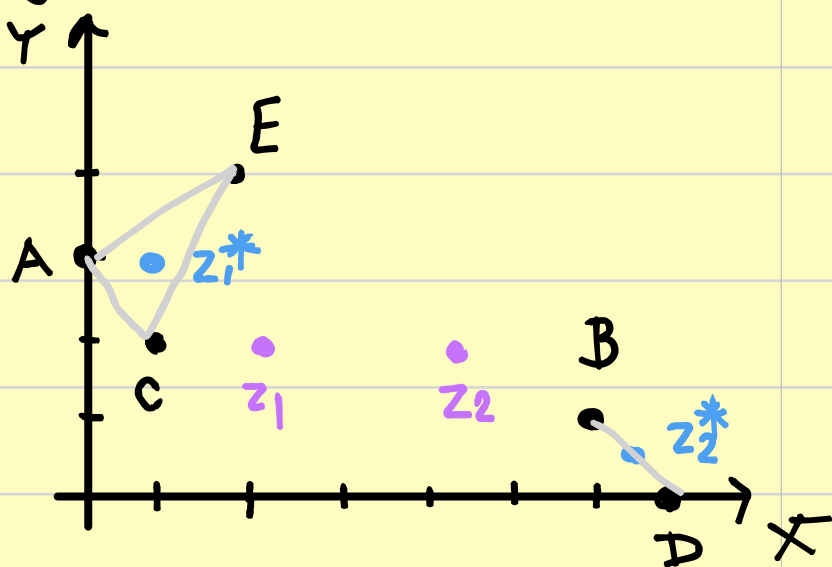
• Schritt 3. Abstand von den Punkten zu den Schwerpunkten

• Schritt 4. Gruppenbildung (Clustern) nach dem geringsten Abstand und ggf. bei Schritt 1. anfangen bis der Abstand zu den Schwerpunkten konstant ist.

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

$$X: [0, 6, 1, 7, 2] \quad Y: [3, 1, 2, 0, 4]$$

A B C D E A B C D E



Bitte fangen Sie mit den Gruppen:

Sch. 1 $G_1[A, B, C]$ $G_2[D, E]$

Sch. 2 Schwerpunkte: [Zentroide]

$$z_1 = \left[\frac{0+6+1}{3}, \frac{3+1+2}{3} \right] = [2.33, 2] ; z_2 = \left[\frac{7+2}{2}, \frac{0+4}{2} \right] = [4.5, 2]$$

Sch. 3 Abstände der Punkte zum Zentroid

$$d_{A,z_1} = \sqrt{(0-2.33)^2 + (3-2)^2} = 2.535 ; d_{A,z_2} = \sqrt{(0-4.5)^2 + (3-2)^2} = 4.609$$

$$d_{B,z_1} = \sqrt{(6-2.33)^2 + (1-2)^2} = 3.804 ; d_{B,z_2} = \sqrt{(6-4.5)^2 + (1-2)^2} = 1.803$$

$$d_{C,z_1} = \sqrt{(1-2.33)^2 + (2-2)^2} = 1.33 ; d_{C,z_2} = \sqrt{(1-4.5)^2 + (2-2)^2} = 3.5$$

$$d_{D,z_1} = \sqrt{(7-2.33)^2 + (0-2)^2} = 5.08 ; d_{D,z_2} = \sqrt{(7-4.5)^2 + (0-2)^2} = 3.2$$

$$d_{E,z_1} = \sqrt{(2-2.33)^2 + (4-2)^2} = 2.027 ; d_{E,z_2} = \sqrt{(2-4.5)^2 + (4-2)^2} = 3.2$$

Sch. 4. Neue Gruppen: $G_1^* = [A, C, E]$ $G_2^* = [B, D]$

Schritt 1*

$$X: [0, 6, 1, 7, 2] \quad Y: [3, 1, 2, 0, 4]$$

A B C D E A B C D E

Schritt 2*. Zentroide.

$$z_1^* = \left[\frac{0+1+2}{3}, \frac{3+2+4}{3} \right] = [1, 3] ; z_2^* = \left[\frac{6+7}{2}, \frac{1+0}{2} \right] = [6.5, 0.5]$$

Schritt 3*. Abstände.

$$\checkmark d_{A,z_1^*} = \sqrt{(0-1)^2 + (3-3)^2} = 1 ; d_{A,z_2^*} = \sqrt{(0-6.5)^2 + (3-0.5)^2} = 6.9$$

$$\checkmark d_{C,z_1^*} = \sqrt{(1-1)^2 + (2-3)^2} = 1 ; d_{C,z_2^*} = \sqrt{(1-6.5)^2 + (2-0.5)^2} = 5.7$$

$$\checkmark d_{E,z_1^*} = \sqrt{(2-1)^2 + (4-3)^2} = 1.414 ; d_{E,z_2^*} = \sqrt{(2-6.5)^2 + (4-0.5)^2} = 5.7$$

$$\checkmark d_{B,z_1^*} = \sqrt{(6-1)^2 + (1-3)^2} = 5.38 ; d_{B,z_2^*} = \sqrt{(6-6.5)^2 + (1-0.5)^2} = 0.707$$

$$\checkmark d_{D,z_1^*} = \sqrt{(7-1)^2 + (0-3)^2} = 6.7 ; d_{D,z_2^*} = \sqrt{(7-6.5)^2 + (0-0.5)^2} = 0.707$$

Position der Lager: $z_1^* [1, 3] \quad z_2^* [6.5, 0.5]$
 Gruppen: $G_1^* [A, C, E] \quad G_2^* [B, D]$

Beispiel. Die Positionen von 6 Werken mit unterschiedlichen Bedarfen an Rohware 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 Algorithmus um der Geschäftsführung eine Empfehlung für die Lagerpositionen auszusprechen.

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

$Y[3, 2, 1, 1, 2, 3]$

Bedarfe $[2, 1, 3, 1, 3, 1]$

Stichwort: gewichtete Mittelwertermittlung.