

# Aufgabe 21: k-Means per Hand

Loyds - Algorithmus:

1) Wähle k Start - Clusterzentren  $\vec{c}_n^{(0)}$

2) Ordne die Daten nach Abstand  $|\vec{c}_n^{(i)} - \vec{x}_m|$

zu den k Clustern zu

3) Berechne die neuen Clusterzentren  
als den Schwerpunkt des Clusters

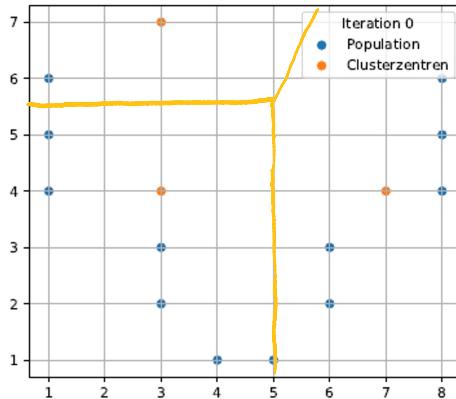
$$\vec{c}_n^{(i+1)} = \frac{1}{N_n} \sum_{m=0}^{N_n} \vec{x}_m$$

4) nächste Iteration startet bei 2)

Population: (1;4) (1;5) (1;6) (3;3) (3;2) (4;1) (5;1) (6;2) (6;3) (8;4) (8;5) (8;6)

a)

1)



Zuordnung zu Clustern: (bei (5,1) ist der Abstand gleich)

$$\vec{c}_1 = (3, 7) : (1, 6)$$

$$\vec{c}_1 = (3, 7) : (1, 6)$$

$$\vec{c}_2 = (3, 4) : (1, 5); (1, 4); (3, 3); (3, 2); (4, 1); (5, 1)$$

$$\vec{c}_3 = (7, 4) : (6, 2); (6, 3); (8, 4); (8, 5); (8, 6)$$

neue Clusterzentren:

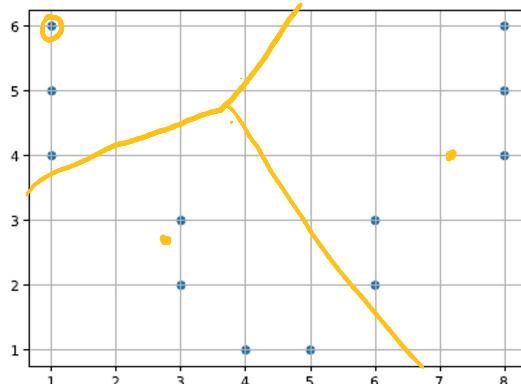
$$\vec{c}_1 = (1, 6)$$

$$\vec{c}_2 = \left( \frac{17}{6}, \frac{8}{3} \right) = (2.8\bar{3}, 2.\bar{6})$$

$$\vec{c}_3 = \left( \frac{36}{5}, 4 \right) = (7.2, 4)$$

b)

2)



Zuordnung zu Clustern:

$$\vec{c}_1 = (1, 6) : (1, 6); (1, 5); (1, 4)$$

$$\vec{c}_2 = (2.8\bar{3}, 2.\bar{6}) : (3, 3); (3, 2); (4, 1); (5, 1)$$

$$\vec{c}_3 = (7.2, 4) : (6, 2); (6, 3); (8, 4); (8, 5); (8, 6)$$

Abstände die nicht klar abzulesen sind:

$$|(1,4) - \vec{c}_1| \approx 2$$

$$|(1,4) - \vec{c}_1| = 2$$

$$|(1,4) - \vec{c}_2| = 3,8$$

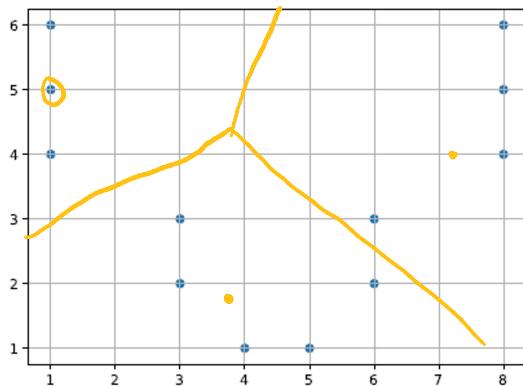
neue Clusterzentren:

$$\vec{c}_1 = (1, 5)$$

$$\vec{c}_2 = \left(\frac{15}{4}, \frac{7}{4}\right) = (3.75, 1.75)$$

$$\vec{c}_3 = \left(\frac{26}{5}, 4\right) = (7.2, 4)$$

3)



Zuordnung zu Clustern:

$$\vec{c}_1 = (1, 5) : (1, 1); (1, 2); (1, 3); (1, 4)$$

$$\vec{c}_2 = (3.75, 1.75) : (3, 3); (3, 2); (4, 1); (5, 1); (6, 2)$$

$$\vec{c}_3 = (7.2, 4) : (6, 3); (8, 4); (8, 5); (8, 6)$$

Abstände die nicht klar abzulesen sind:

$$|(6,2) - \vec{c}_2| = 2.264$$

$$|(6,2) - \vec{c}_3| = 2.332$$

neue Clusterzentren:

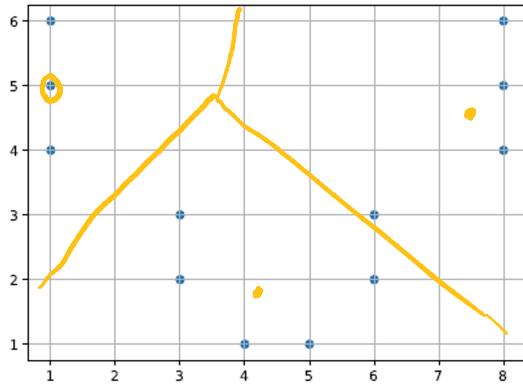
neue Clusterzentren:

$$\vec{c}_1 = (1, 5)$$

$$\vec{c}_2 = \left(\frac{21}{5}, \frac{9}{5}\right) = (4.2, 1.8)$$

$$\vec{c}_3 = \left(\frac{15}{2}, \frac{9}{2}\right) = (7.5, 4.5)$$

4)



Zuordnung zu Clustern:

$$\vec{c}_1 = (1, 5) : (1, 4); (1, 5); (1, 6)$$

$$\vec{c}_2 = (4.2, 1.8) : (3, 3); (3, 2); (4, 1); (5, 1); (6, 2)$$

$$\vec{c}_3 = (7.5, 4.5) : (6, 3); (8, 4); (8, 5); (8, 6)$$

Abstände die nicht klar abzuksen sind:

$$|(6, 3) - \vec{c}_2| = 2.163$$

$$|(6, 3) - \vec{c}_3| = 2.121$$

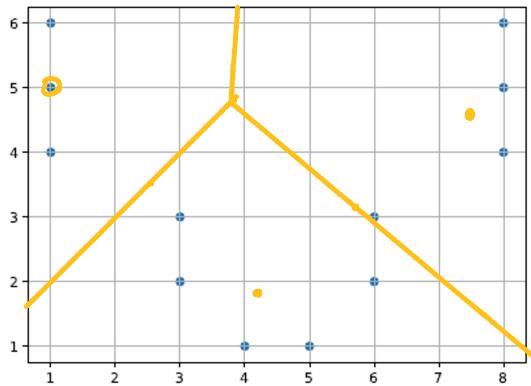
An den Clusterzuordnungen hat sich nichts geändert.

Dadurch bleiben auch die Clusterzentren gleich.

Also ist der Algorithmus in ein Gleichgewicht konvergiert.

Finale Zuordnung:

$$\vec{c}_1 = (1, 5); \vec{c}_2 = (4.2, 1.8); \vec{c}_3 = (7.5, 4.5)$$



c)

Nach 3 bzw. 4 Iterationen ist der Algorithmus konvergiert.

Das Ergebnis entspricht nicht unseren Erwartungen,  
denn den Punkt (6, 3) hätten wir zum mittleren  
Cluster zugeordnet.

Man sieht, dass k-Means teilweise stark  
von der Wahl der Initial-Clusterzentren abhängt.

Unsere Einschätzung / Erwartung:

$$\vec{c}_1 = (1, 5); \vec{c}_2 = (4.5, 2); \vec{c}_3 = (8, 5)$$

