# K-means Beispiel mit $k = 3$

Angenommen, wir haben die folgenden fünf Datenpunkte in einem zweidimensionalen Raum:

$$\text{Punkte} = \{(2,3), (3,3), (6,7), (8,8), (10,10)\}$$

Wir möchten diese Punkte in $k = 3$ Cluster gruppieren.

## Schritt 1: Initialisierung der Zentroiden

Wir wählen zunächst zufällig drei Datenpunkte als Anfangszentroiden:

- Zentroid 1: $C_1 = (2,3)$
- Zentroid 2: $C_2 = (6,7)$
- Zentroid 3: $C_3 = (8,8)$

## Schritt 2: Berechnung der Zuordnung

Nun berechnen wir die euklidischen Distanzen jedes Datenpunkts zu den drei Zentroiden und weisen jeden Punkt dem nächstgelegenen Zentroiden zu.

Die euklidische Distanz zwischen zwei Punkten $(x_1, y_1)$ und $(x_2, y_2)$ ist gegeben durch:

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

**Punkt (2, 3):**

$$\begin{aligned}
\text{Distanz zu } C_1 = (2,3) &: \quad \sqrt{(2-2)^2 + (3-3)^2} = 0 \\
\text{Distanz zu } C_2 = (6,7) &: \quad \sqrt{(6-2)^2 + (7-3)^2} = \sqrt{16+16} = \sqrt{32} \approx 5{,}66 \\
\text{Distanz zu } C_3 = (8,8) &: \quad \sqrt{(8-2)^2 + (8-3)^2} = \sqrt{36+25} = \sqrt{61} \approx 7{,}81
\end{aligned}$$

Nächstgelegener Zentroid: $C_1$

**Punkt (3, 3):**

$$\begin{aligned}
\text{Distanz zu } C_1 = (2,3) &: \quad \sqrt{(3-2)^2 + (3-3)^2} = \sqrt{1} = 1 \\
\text{Distanz zu } C_2 = (6,7) &: \quad \sqrt{(6-3)^2 + (7-3)^2} = \sqrt{9+16} = \sqrt{25} = 5 \\
\text{Distanz zu } C_3 = (8,8) &: \quad \sqrt{(8-3)^2 + (8-3)^2} = \sqrt{25+25} = \sqrt{50} \approx 7{,}07
\end{aligned}$$

Nächstgelegener Zentroid: $C_1$

**Punkt (6, 7):**

$$\begin{aligned}
\text{Distanz zu } C_1 = (2,3) &: \quad \sqrt{(6-2)^2 + (7-3)^2} = \sqrt{16+16} = \sqrt{32} \approx 5{,}66 \\
\text{Distanz zu } C_2 = (6,7) &: \quad \sqrt{(6-6)^2 + (7-7)^2} = 0 \\
\text{Distanz zu } C_3 = (8,8) &: \quad \sqrt{(8-6)^2 + (8-7)^2} = \sqrt{4+1} = \sqrt{5} \approx 2{,}24
\end{aligned}$$

Nächstgelegener Zentroid: $C_2$

**Punkt (8, 8):**

$$\begin{aligned}
\text{Distanz zu } C_1 = (2,3): &\quad \sqrt{(8-2)^2+(8-3)^2} = \sqrt{36+25} = \sqrt{61} \approx 7{,}81 \\
\text{Distanz zu } C_2 = (6,7): &\quad \sqrt{(8-6)^2+(8-7)^2} = \sqrt{4+1} = \sqrt{5} \approx 2{,}24 \\
\text{Distanz zu } C_3 = (8,8): &\quad \sqrt{(8-8)^2+(8-8)^2} = 0
\end{aligned}$$

Nächstgelegener Zentroid: $C_3$

**Punkt (10, 10):**

$$\begin{aligned}
\text{Distanz zu } C_1 = (2,3): &\quad \sqrt{(10-2)^2+(10-3)^2} = \sqrt{64+49} = \sqrt{113} \approx 10{,}63 \\
\text{Distanz zu } C_2 = (6,7): &\quad \sqrt{(10-6)^2+(10-7)^2} = \sqrt{16+9} = \sqrt{25} = 5 \\
\text{Distanz zu } C_3 = (8,8): &\quad \sqrt{(10-8)^2+(10-8)^2} = \sqrt{4+4} = \sqrt{8} \approx 2{,}83
\end{aligned}$$

Nächstgelegener Zentroid: $C_3$

## Zuordnung nach dem ersten Schritt

- $C_1$: {(2, 3), (3, 3)}
- $C_2$: {(6, 7)}
- $C_3$: {(8, 8), (10, 10)}

### Schritt 3: Zentroiden aktualisieren

Wir berechnen die neuen Zentroiden, indem wir den Mittelwert der Punkte in jedem Cluster berechnen.

- Neuer Zentroid für $C_1$: Mittelwert von {(2, 3), (3, 3)} $= \left(\frac{2+3}{2}, \frac{3+3}{2}\right) = (2.5, 3)$
- Neuer Zentroid für $C_2$: Da es nur einen Punkt gibt, bleibt der Zentroid gleich $(6, 7)$.
- Neuer Zentroid für $C_3$: Mittelwert von {(8, 8), (10, 10)} $= \left(\frac{8+10}{2}, \frac{8+10}{2}\right) = (9, 9)$

### Schritt 4: Wiederholung

Wir würden nun die Schritte 2 und 3 wiederholen, indem wir die Zuordnung der Punkte basierend auf den aktualisierten Zentroiden erneut berechnen und die Zentroiden so lange aktualisieren, bis sich die Zuordnungen nicht mehr ändern.

### Endergebnis

Nach mehreren Iterationen (in diesem einfachen Beispiel könnten bereits 2 Iterationen ausreichen) würden wir eine stabile Zuordnung und endgültige Zentroiden haben. Das Ergebnis wären die 3 Cluster mit den zugehörigen Punkten.