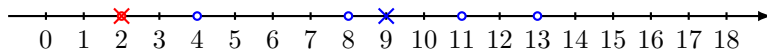
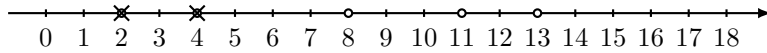


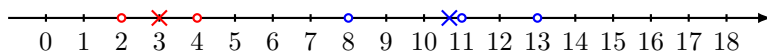
Aufgabe 1

- Aus den Datenpunkten werden zufällig k Clusterzentren ausgewählt.
- Jeder Datenpunkt wird zufällig einem der k Cluster zugeordnet.

Aufgabe 2

$$C_1 = 2/1 = 2 \quad C_2 = (4 + 8 + 11 + 13)/4 = 9$$

$$J = 0^2 + 5^2 + 1^2 + 2^2 + 4^2 = 46$$



$$C_1 = (2 + 4)/2 = 3 \quad C_2 = (8 + 11 + 13)/3 = 10\frac{2}{3}$$

$$J = 1^2 + 1^2 + \left(\frac{8}{3}\right)^2 + \left(\frac{1}{3}\right)^2 + \left(\frac{7}{3}\right)^2 = \frac{3+3+64+1+49}{3} = \frac{120}{3} = 40$$

J ist die Summe der quadrierten Abstände aller Punkte von ihrem jeweiligen Clusterzentrum.

Aufgabe 3

Nein, der k -Means-Algorithmus findet im Allgemeinen ein *lokales Optimum*. Um das Ergebnis zu verbessern, kann man das Verfahren mehrfach mit jeweils anderen Startwerten durchführen, und dann das Resultat mit dem kleinsten Distortion Measure J wählen.

Aufgabe 4

- Markt- und Kundensegmentierung
- Bildsegmentierung
- Astronomie (verwendet Eigenschaften von Gammablitzern um die zugrundeliegenden Ereignisse zu typisieren)