

K-Nearest Neighbours (KNN)

- Wir hinzufügen neue Daten in bereits bestehenden Klassen zu integrieren.

Problemstellung:

Sendung	Gewicht (kg)	Volumen (m³)	Klasse
A	10	0'02	N
B	50	0'15	S
C	30	0'10	N
D	60	0'20	S
E	20	0'05	N

X	40	0'12	?
---	----	------	---

1. Schritt Normierung

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

A $\frac{10-10}{60-10} = 0$ $\frac{0'02-0'02}{0'2-0'02} = 0$ N

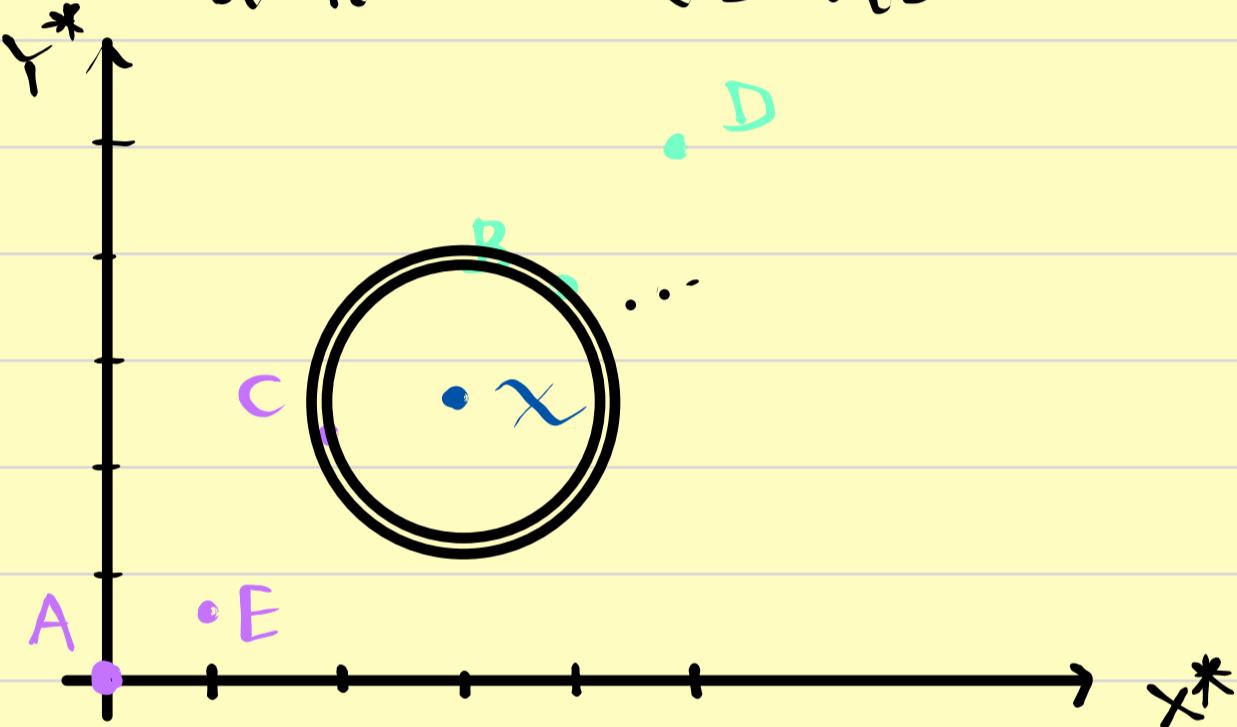
B $\frac{50-10}{60-10} = 0'8$ $\frac{0'15-0'02}{0'2-0'02} = 0'72$ S

$$C \quad \frac{30-10}{60-10} = 0.4 \quad \frac{0.1-0.02}{0.2-0.02} = 0.44 \text{ N}$$

$$D \quad \frac{60-10}{60-10} = 1 \quad \frac{0.2-0.02}{0.2-0.02} = 1 \text{ s}$$

$$E \quad \frac{20-10}{60-10} = 0.2 \quad \frac{0.05-0.02}{0.2-0.02} = 0.167 \text{ N}$$

$$X \quad \frac{40-10}{60-10} = 0.6 \quad \frac{0.12-0.02}{0.2-0.02} = 0.55 \text{ ?}$$



2. Schritt

$$d_{A,x} = \sqrt{(0-0.6)^2 + (0-0.55)^2} = 0.8139 \text{ N}$$

$$d_{B,x} = \sqrt{(0.5-0.6)^2 + (0.72-0.55)^2} = 0.26 \text{ s}$$

$$d_{C,x} = \sqrt{(0.4-0.6)^2 + (0.44-0.55)^2} = 0.23 \text{ N}$$

$$d_{D,x} = \sqrt{(1-0.6)^2 + (1-0.55)^2} = 0.602 \text{ s}$$

$$d_{E,X} = \sqrt{(1-t'^2)^2 + (1-t'^1_{167})^2} = 0.5537 \text{ N}$$

3. Schritt.

$$d_{CX} < d_{BX} < d_{EX} < d_{DX} < d_{AX}$$



4. Schritt.

Ich nehme die N, N und schau die W dafür, dass der neue Punkt zu den Klassen gehört.

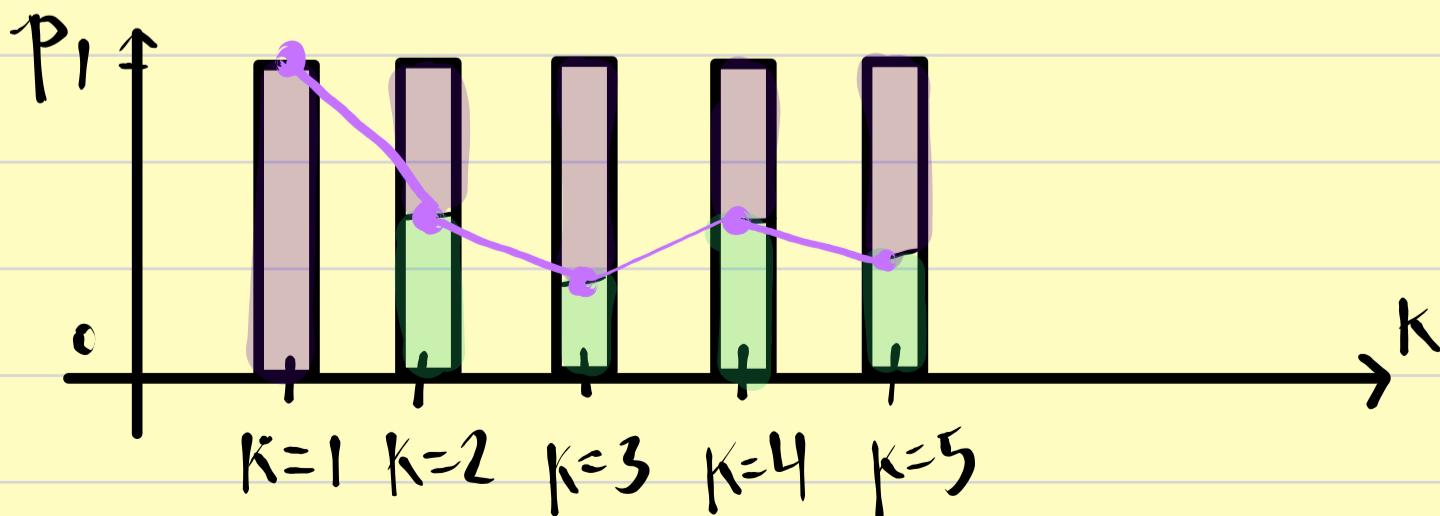
$$k=1 : p(X \in N) = 1 ; p(X \in S) = 0$$

$$k=2 : p(X \in N) = 0.5 ; p(X \in S) = 0.5$$

$$k=3 : p(X \in N) = \frac{2}{3} = 0.66 ; p(X \in S) = 0.33$$

$$k=4 : p(X \in N) = 0.5 ; p(X \in S) = 0.5$$

$$k=5 : p(X \in N) = \frac{3}{5} = 0.6 ; p(X \in S) = 0.4$$



Bei $K=3$ trennen sich die Daten am besten.

$K=1$: nicht ausreichende Daten

$K=2$: Klasse $N = 0^15$ | $\Delta = 0 \rightarrow$ keine Trennung.
Klasse $S = 0^15$

$K=3$: Klasse $N = 0^166$ | $\Delta = 0^133 \rightarrow \text{OK}$
Klasse $S = 0^133$

$K=4$: analog $K=2$.

$K=5$: Klasse $N = 0^16$ | $\Delta = 0^12 \rightarrow \text{X}$
Klasse $S = 0^14$

Der Punkt X gehört zu Klasse N .

* Shannon Informationstheorie.

Übung: Gegeben ist die Position von 6 Werten mit $[x, y]$ Koordinaten.

1) Bitte mit K-Means Clustering die Gruppen in 2 Kategorien teilen.

2) Bitte mit KNN entscheiden, zu welchen Gruppe Wert α gehört?

$$w \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix}$$
$$x \begin{bmatrix} 0, 0, 1, 3, 4, 4 \end{bmatrix} \begin{bmatrix} 2 \end{bmatrix}$$
$$y \begin{bmatrix} 0, 2, 1, 2, 1, 3 \end{bmatrix} \begin{bmatrix} 5 \end{bmatrix}$$