

# K-Nearest Neighbors (KNN)

Wir bezwecken neue Daten in bereits bestehenden Klassen zu integrieren.

Problemstellung:

Sending	Gewicht (kg)	Volumen (m <sup>3</sup> )	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

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

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

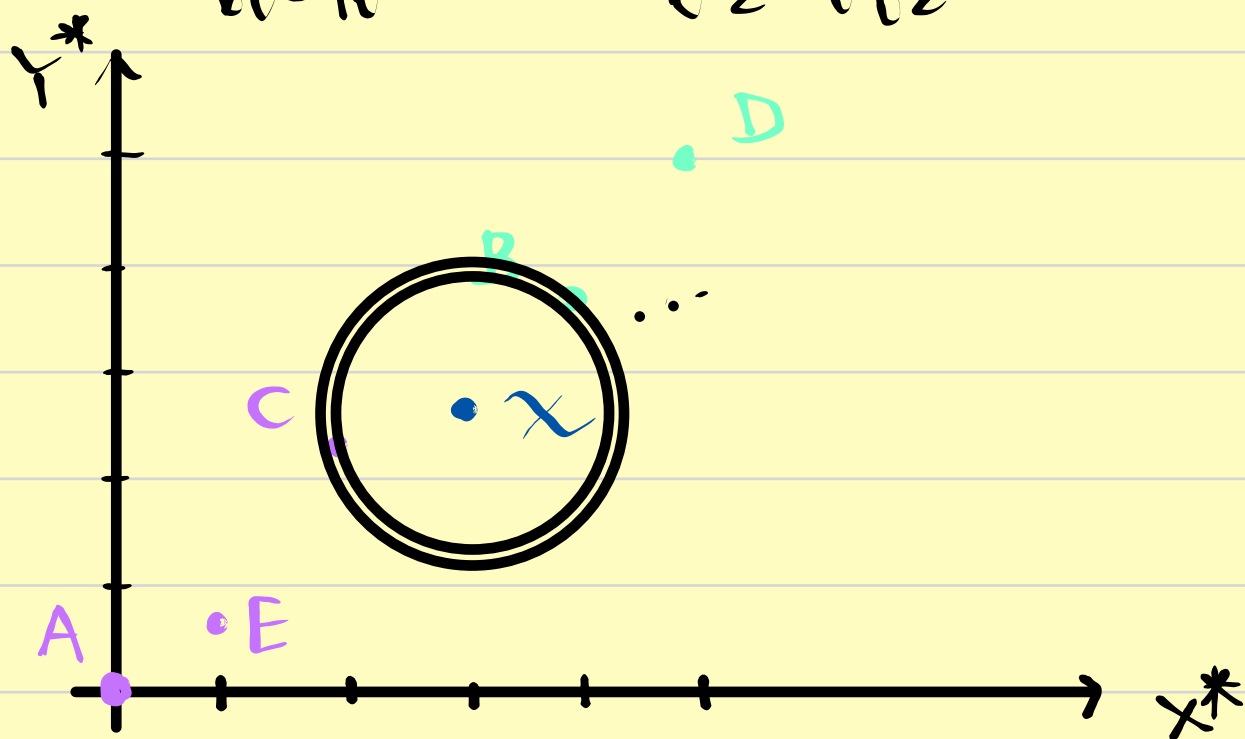
B  $\frac{50-10}{60-10} = 0'8$   $\frac{0'15-0'02}{0'2-0'02} = 0'72$  S

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

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

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

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



## 2. Schritt

$$d_{A,X} = \sqrt{(0-0.6)^2 + (1-0.55)^2} = 0.8139 \quad N$$

$$d_{B,X} = \sqrt{(0.8-0.6)^2 + (0.72-0.55)^2} = 0.26 \quad S$$

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

$$d_{D,X} = \sqrt{(1-0.6)^2 + (1-0.55)^2} = 0.602 \quad S$$

$$d_{E,X} = \sqrt{(1-0'2)^2 + (1-0'167)^2} = 0'5537 \text{ N}$$

3. Schritt.

$$d_{C,X} < d_{B,X} < d_{E,X} < d_{D,X} < d_{A,X}$$

N
S
N
S
N

4. Schritt.

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

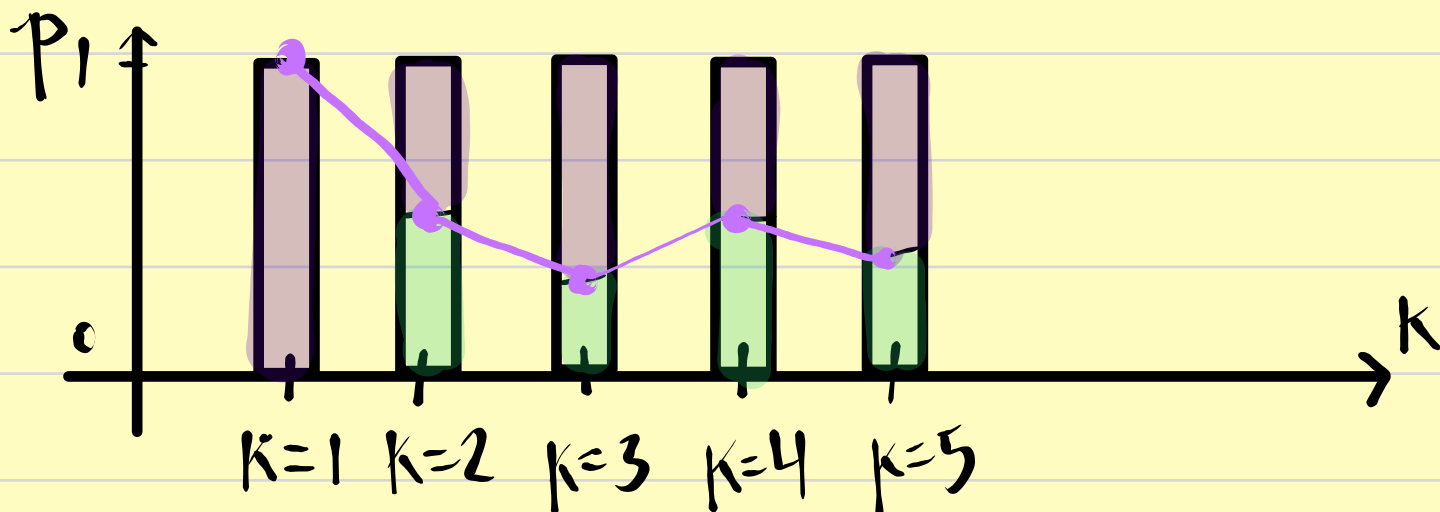
$$k=1 : p(X \in N) = 1 ; p(X \in S) = 0$$

$$k=2 : p(X \in N) = 0'5 ; p(X \in S) = 0'5$$

$$k=3 : p(X \in N) = \frac{2}{3} = 0'66 ; p(X \in S) = 0'33$$

$$k=4 : p(X \in N) = 0'5 ; p(X \in S) = 0'5$$

$$k=5 : p(X \in N) = \frac{3}{5} = 0'6 ; p(X \in S) = 0'4$$



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

$k=1$ : nicht ausreichende Daten

$k=2$ : Klasse  $N=0'5$  |  $\Delta = 0 \rightarrow$  keine Trennung.  
Klasse  $S=0'5$

$k=3$ : Klasse  $N=0'66$  |  $\Delta = 0'33 \rightarrow \checkmark$   
Klasse  $S=0'33$

$k=4$ : analog  $k=2$ .

$k=5$ : Klasse  $N=0'6$  |  $\Delta = 0'2 \rightarrow \otimes$   
Klasse  $S=0'4$

Der Punkt  $X$  gehört zu Klasse  $N$ .

\* Shannon Informationstheorie.

Uhring: 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 K-MN entscheiden zu welcher Gruppe Wert  $x$  gehört?

$w$	1	2	3	4	5	6	$\lambda$
$x$	[0	, 0	, 1	, 3	, 4	, 4]	[2]
$y$	[0	, 2	, 1	, 2	, 1	, 3]	[5]