

K-NEAREST NEIGHBOUR (K-NN)

K-Neans Clustering
(1. Semester)

Problemstellung.

Sendung	Gewicht (kg)	Volumen (m ³)	Kategorie
A	10	0'02	Normal
B	50	0'15	Schwer
C	30	0'10	Normal
D	60	0'20	Schwer
E	20	0'05	Normal
X	40	0'12	<div style="border: 1px solid black; padding: 5px; display: inline-block;">?</div>

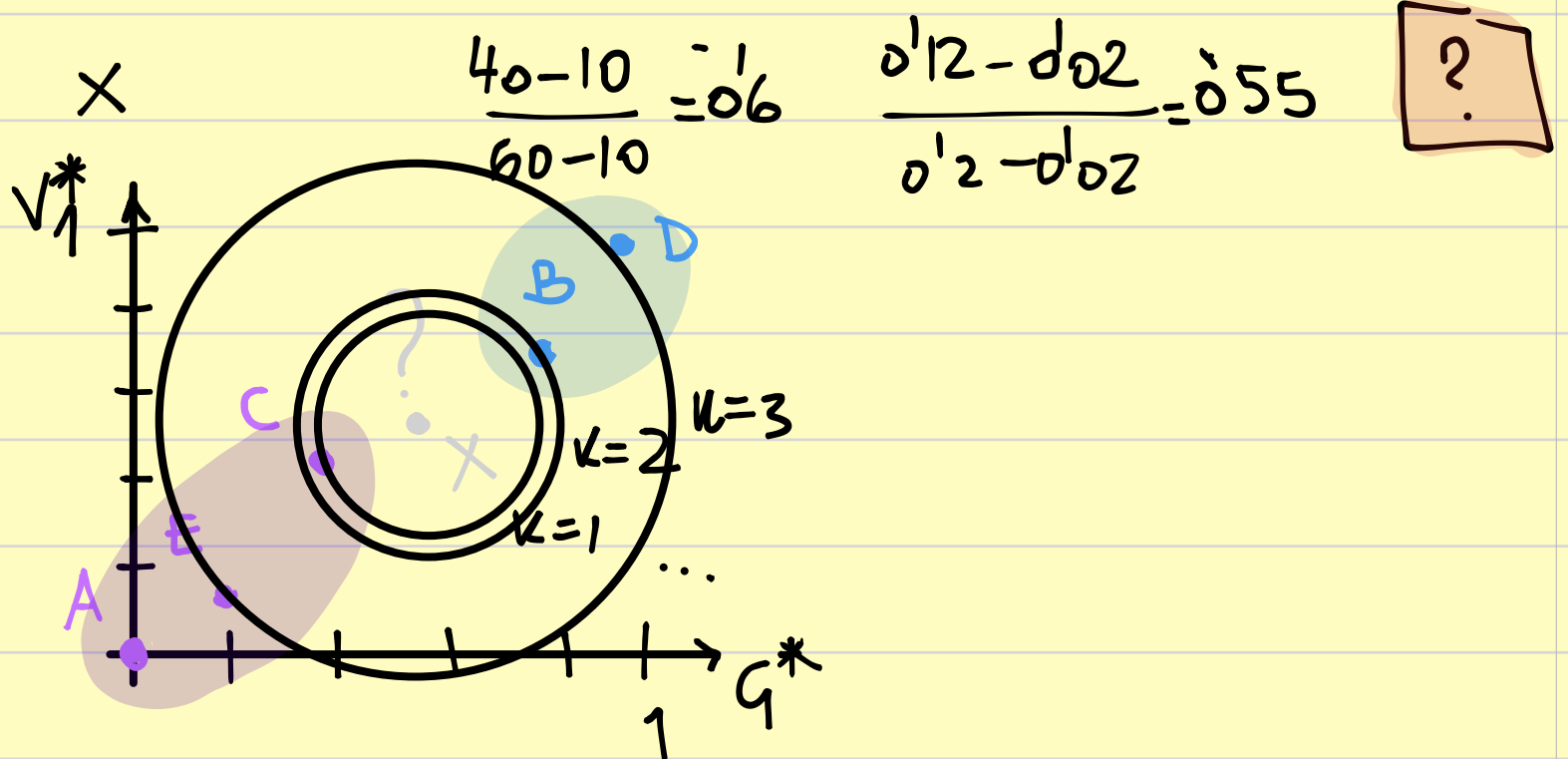
Anhand bereits geklusterter Daten, sind wir in der Lage neue Daten einer Klasse zuzuordnen.

VORTEIL. wir benötigen nicht eine rekursive Herangehensweise (wie beim K-Means Clustering), KNN ist ein s.g. "Lazy Learning" Algorithmus.

SCHRITTE K-NN.

1. SCHRITT. NORMIEREN. $x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$
(Min-max Skalierung)

Sendung	Gewicht*	Volumen*	Kategorie
A	$\frac{10-10}{60-10} = 0$	$\frac{0'02-0'02}{0'2-0'02} = 0$	Normal
B	$\frac{50-10}{60-10} = 0'8$	$\frac{0'15-0'02}{0'2-0'02} = 0'72$	Schwer
C	$\frac{30-10}{60-10} = 0'4$	$\frac{0'10-0'02}{0'2-0'02} = 0'44$	Normal
D	$\frac{60-10}{60-10} = 1$	$\frac{0'2-0'02}{0'2-0'02} = 1$	Schwer
E	$\frac{20-10}{60-10} = 0'2$	$\frac{0'05-0'02}{0'2-0'02} = 0'167$	Normal



2. SCHRITT. Abstände vom neuen Punkt zu allen Anderen.

$$\begin{aligned}
 d_{xA} &= \sqrt{(0-0'6)^2 + (0-0'55)^2} = 0'8139 \quad \text{N} \\
 d_{xB} &= \sqrt{(0'8-0'6)^2 + (0'72-0'55)^2} = 0'2625 \quad \text{S} \\
 d_{xC} &= \sqrt{(0'4-0'6)^2 + (0'44-0'55)^2} = 0'228 \quad \text{N} \\
 d_{xD} &= \sqrt{(1-0'6)^2 + (1-0'55)^2} = 0'602 \quad \text{S} \\
 d_{xE} &= \sqrt{(0'2-0'6)^2 + (0'167-0'55)^2} = 0'5537 \quad \text{N}
 \end{aligned}$$

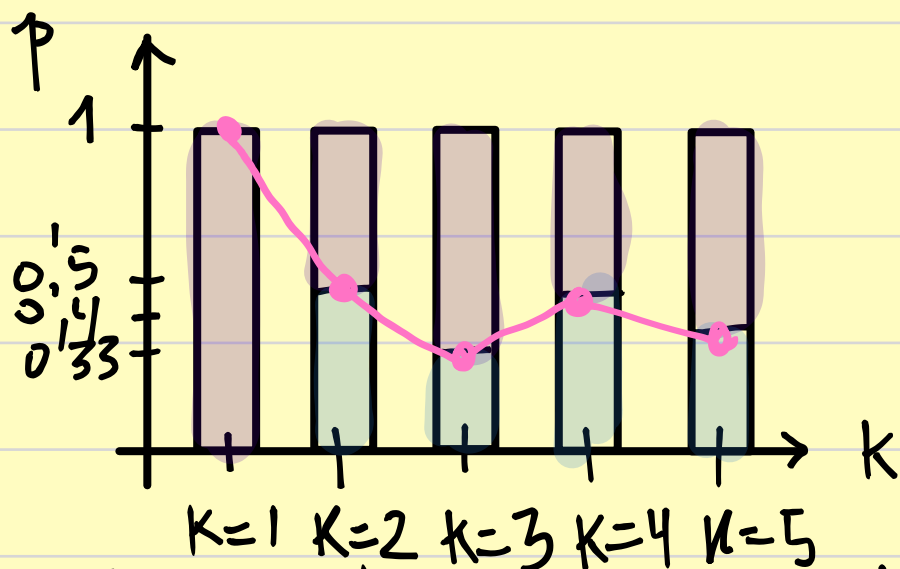
Sendung	Gewicht*	Volumen*	Kategorie
A	$\frac{10-10}{60-10} = 0$	$\frac{0'02-0'02}{0'2-0'02} = 0$	Normal
B	$\frac{50-10}{60-10} = 0'8$	$\frac{0'15-0'02}{0'2-0'02} = 0'72$	Schwer
C	$\frac{30-10}{60-10} = 0'4$	$\frac{0'10-0'02}{0'2-0'02} = 0'44$	Normal
D	$\frac{60-10}{60-10} = 1$	$\frac{0'2-0'02}{0'2-0'02} = 1$	Schwer
E	$\frac{20-10}{60-10} = 0'2$	$\frac{0'05-0'02}{0'2-0'02} = 0'167$	Normal
X	$\frac{40-10}{60-10} = 0'6$	$\frac{0'12-0'02}{0'2-0'02} = 0'55$?

3. SCHRITT. Ordnen der Abstände aufsteigend.

$$d_{xC} < d_{xB} < d_{xE} < d_{xD} < d_{xA}$$

N
S
N
S
N

4. SCHRITT. Mehrheitsentscheidung.



k=1 bedeutet, wir nehmen nur einen Nachbar; die W. dafür dass x zur Kategorie N gehört ist 100%.

k=2 bedeutet, wir nehmen zwei Nachbar; die W. dafür, dass x zur Kategorie N gehört ist 50%.

$K=3$ bedeutet, wir nehmen drei Nachbar, die W. dafür, dass X zur Kategorie N gehört ist 66%.

...

4. SCHRITT. Entscheidung der K .

Anhand der Wahrscheinlichkeiten sehen wir, dass die optimale Verteilung bei $K=3$ erreicht wird.

X gehört also zur Klasse Normal.

Übung.

	X	Y
A	0	0
	0	2
	1	1
B	3	2
	4	1
	4	3

$$d_{xA1} = \sqrt{(2-0)^2 + (5-0)^2} = 5.38$$

$$d_{xA2} = 3.605$$

$$d_{xA3} = 4.123$$

$$d_{xB1} = 3.16$$

$$d_{xB2} = 4.47$$

$$d_{xB3} = 2.82$$

6.

3.

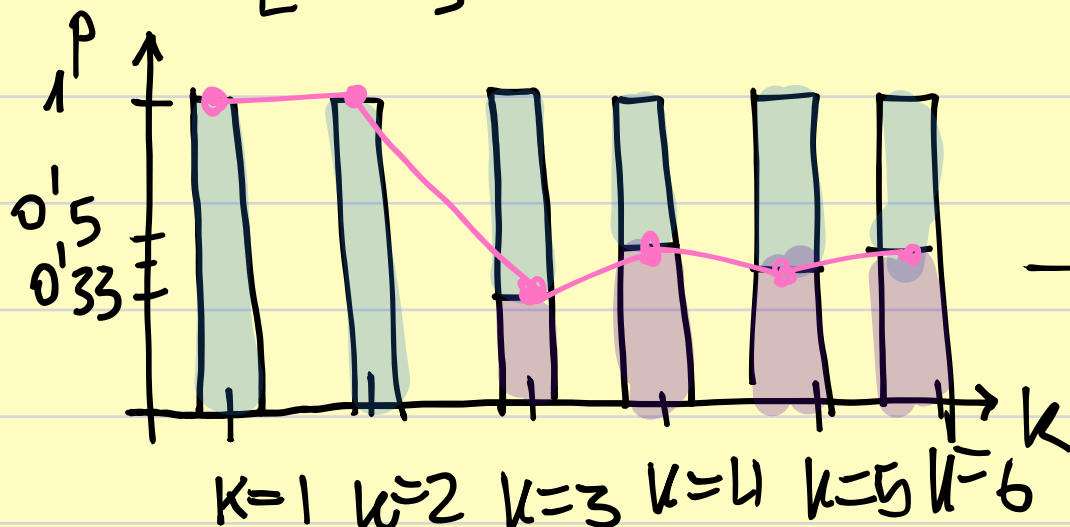
4.

2.

5.

1.

Neuer Kunde $[2, 5]$ Welche Klasse?



→ $K=3$. Klasse B.

