

# K. NEAREST NEIGHBOR (KNN)

Wir bezeichnen neue Daten in bereits bestehenden Klassen (k-Means Clustering) zu integrieren.

. VORTEILE . 1. sehr schnell 2. sehr robust .

. NACHTEILE . 1. vorher müssen die Kategorien bekannt sein.

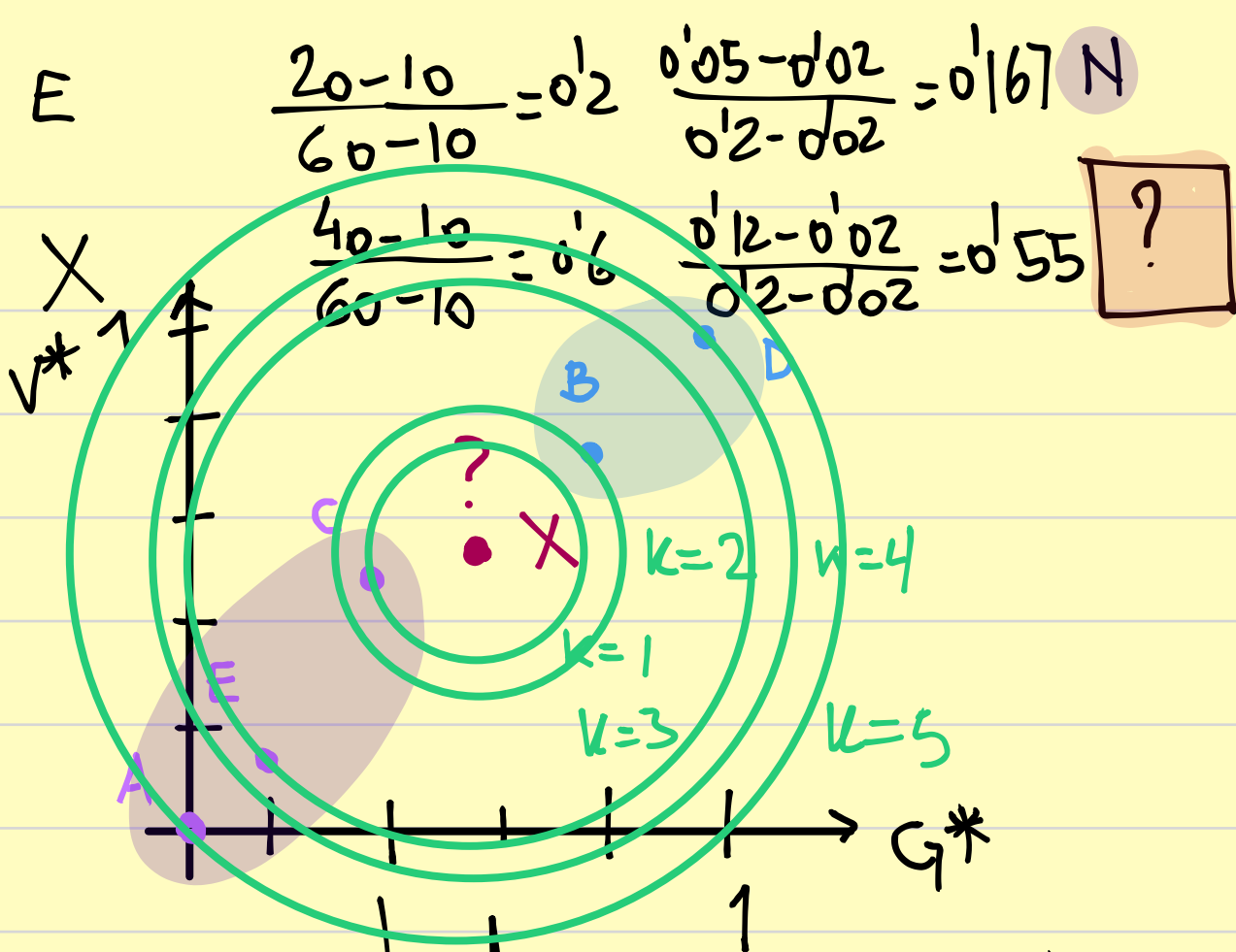
Problemstellung	(kg)	(m <sup>3</sup> )	
Sendung	Gewicht	Volumen	Klasse
A	10	0'02	Normal
B	50	0'15	Schwer
C	30	0'10	Normal
D	60	0'20	Schwer
E	20	0'05	Normal
X	40	0'12	?

KNN ist ein  
..Lazy Learning  
algorithmus



1. SCHRITT NORMIERUNG.  $x_i^* = \frac{x_i - x_{min}}{x_{max} - x_{min}}$  [MinMaxSkal]

Sendung	Gewicht*	Volumen*	Klasse
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	$\frac{30-10}{60-10} = 0'4$	$\frac{0'10-0'02}{0'2-0'02} = 0'44$	N
D	$\frac{60-10}{60-10} = 1$	$\frac{0'2-0'02}{0'2-0'02} = 1$	S



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

$$d_{xA} = \sqrt{(0-0.2)^2 + (0-0.167)^2} = 0.18139 \text{ N}$$

$$d_{xB} = \sqrt{(0.6-0.2)^2 + (0.12-0.167)^2} = 0.2625 \text{ S}$$

$$d_{xC} = \sqrt{(0.4-0.2)^2 + (0.05-0.167)^2} = 0.228 \text{ N}$$

$$d_{xD} = \sqrt{(0.6-0.2)^2 + (0.12-0.167)^2} = 0.2602 \text{ S}$$

$$d_{xE} = \sqrt{(0.2-0.2)^2 + (0.05-0.167)^2} = 0.1167 \text{ N}$$

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

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

4. SCHRITT. Mehrheitsentscheidung

Ich nehme mir  $k$  Nearest Neighbours (Nahliegende Nachbarn) und schaue die W dafür, dass der neue Punkt zu den Klassen gehört.

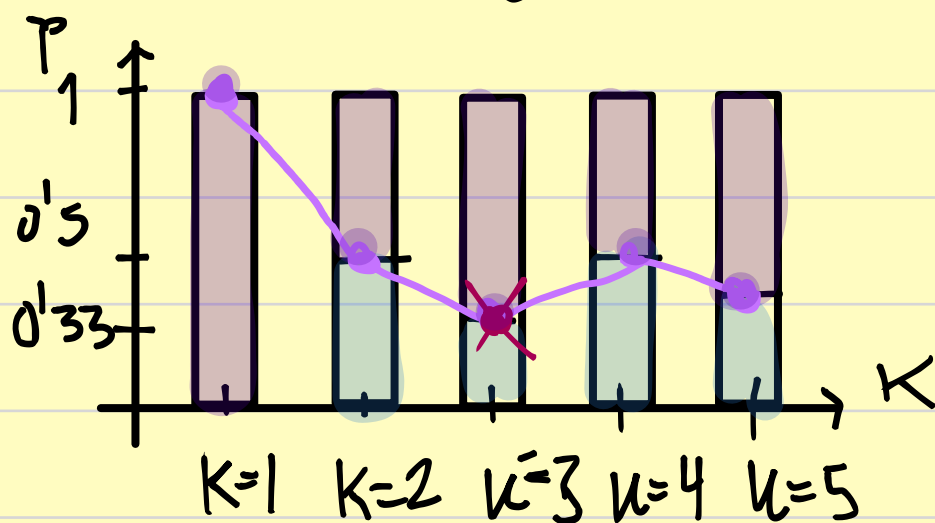
$$k=1 : p(X \in N) = \frac{1}{1} = 1; p(X \in S) = \frac{0}{1} = 0$$

$$k=2 : p(X \in N) = \frac{1}{2} = 0.5; p(X \in S) = \frac{1}{2} = 0.5$$

$$k=3 : p(X \in N) = \frac{2}{3} = 0.66; p(X \in S) = \frac{1}{3} = 0.33$$

$$k=4 : p(X \in N) = \frac{2}{4} = 0.5; p(X \in S) = \frac{2}{4} = 0.5$$

$$k=5 : p(X \in N) = \frac{3}{5} = 0.6; p(X \in S) = \frac{2}{5} = 0.4$$



- Bei  $k=3$  trennen wir die Daten am besten! Min/Max indikatore
- Die Mehrheit der Daten sagt aus, dass bei  $k=3$ , die neue Datei zur klass **N** gehört.

(Shannon Informations Theorie)

Übung. Gegeben ist die Position von 6 Werke mit  $(x, y)$  Koordinaten. 1) Bitte mit k.M.C. die Gruppen in 2 Kategorien teilen 2) Mit kNN entscheiden zu welcher Gruppe Werk  $\alpha$  gehört?

- $w_1[0,0]$   $w_2[0,2]$   $w_3[1,1]$   $w_4[3,2]$   $w_5[4,1]$   $w_6[4,3]$
- $w_\alpha[2,5]$

