

k. Nearest Neighbour (nahliegende Nachbarn)

KNN

Wir bezwecken neue Daten in bereits bestehenden Klassen (Kategorien) [diese kommen von k-Means clustering] zu integrieren.

• VORTEILE: 1) sehr schnell 2) sehr robust

• NACHTEILE: 1) vorher müssen die Kategorien bekannt sein.

k-Means clust.

PROBLEMSTELLUNG:

Sendung	Gewicht (kg)	Volumen (m ³)	Kategorie
A	10	0'02	Normal (N)
B	50	0'15	Schwer (S)
C	30	0'10	N
D	60	0'20	S
E	20	0'05	N
X	40	0'12	?

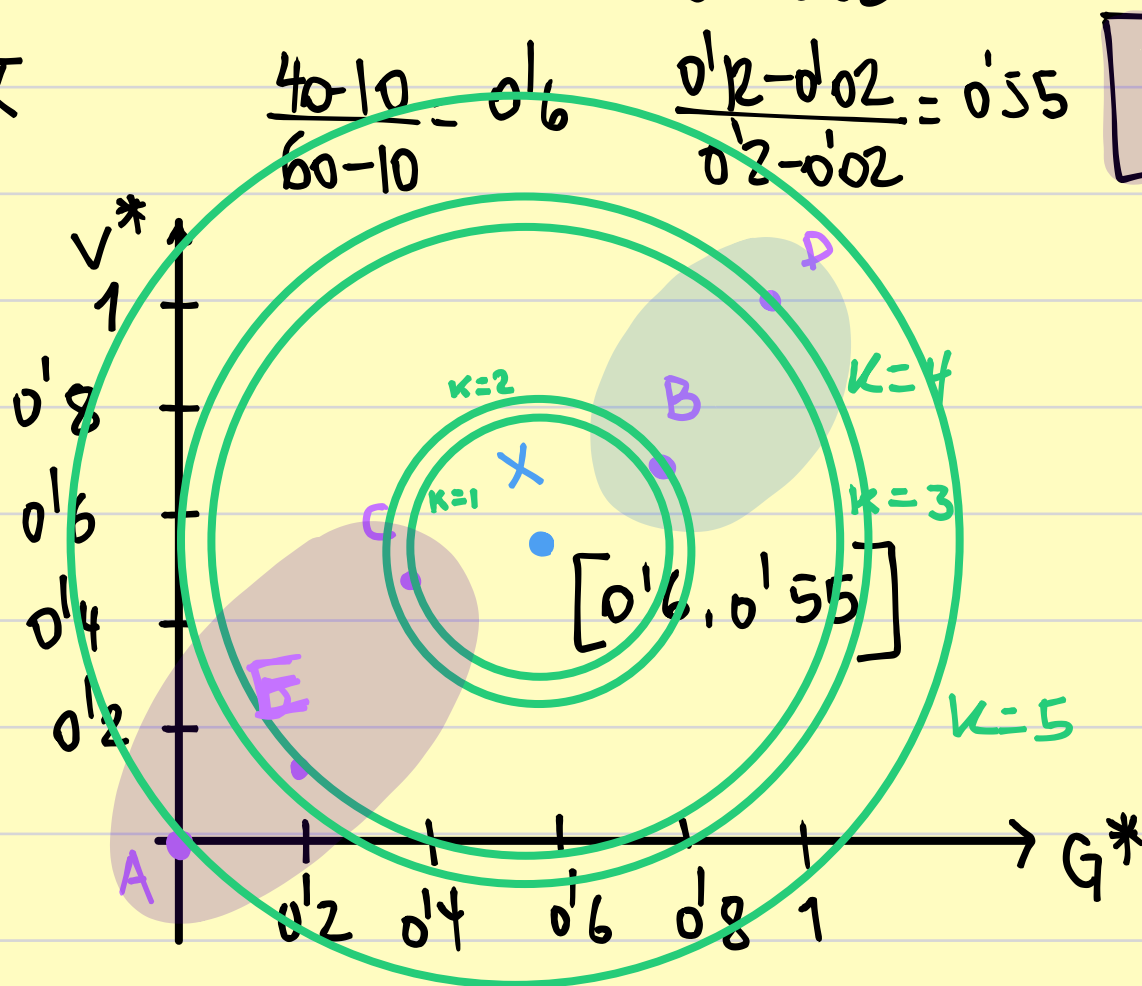
Zu welcher Klasse gehört X?

KNN agiert durch einen „Lazy-Algorithmus“: das Algorithmus lernt nicht. Agiert aus einer datenbasierten Mehrheitsentscheidung.

1. SCHRITT. Normieren. Min-Max Scaler.

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

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	$\frac{30-10}{60-10} = 0'4$	$\frac{0'1-0'02}{0'2-0'02} = 0'44$	N
D	$\frac{60-10}{60-10} = 1$	$\frac{0'2-0'02}{0'2-0'02} = 1$	S
E	$\frac{20-10}{60-10} = 0'2$	$\frac{0'05-0'02}{0'2-0'02} = 0'167$	N
X	$\frac{40-10}{60-10} = 0'6$	$\frac{0'12-0'02}{0'2-0'02} = 0'55$?



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

$$d_{AX} = \sqrt{(0-0'6)^2 + (0-0'55)^2} = 0'8139 \quad N$$

$$d_{BX} = \sqrt{(0'8-0'6)^2 + (0'72-0'55)^2} = 0'262 \quad S$$

$$\Rightarrow d_{CX} = \sqrt{(0'4-0'6)^2 + (0'44-0'55)^2} = 0'228 \quad N$$

$$d_{DX} = \sqrt{(1-0'6)^2 + (1-0'55)^2} = 0'6 \quad S$$

$$d_{EX} = \sqrt{(0'2-0'6)^2 + (0'167-0'55)^2} = 0'55 \quad N$$

3. SCHRITT. Abstand aufsteigend ordnen.

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

N S N S N

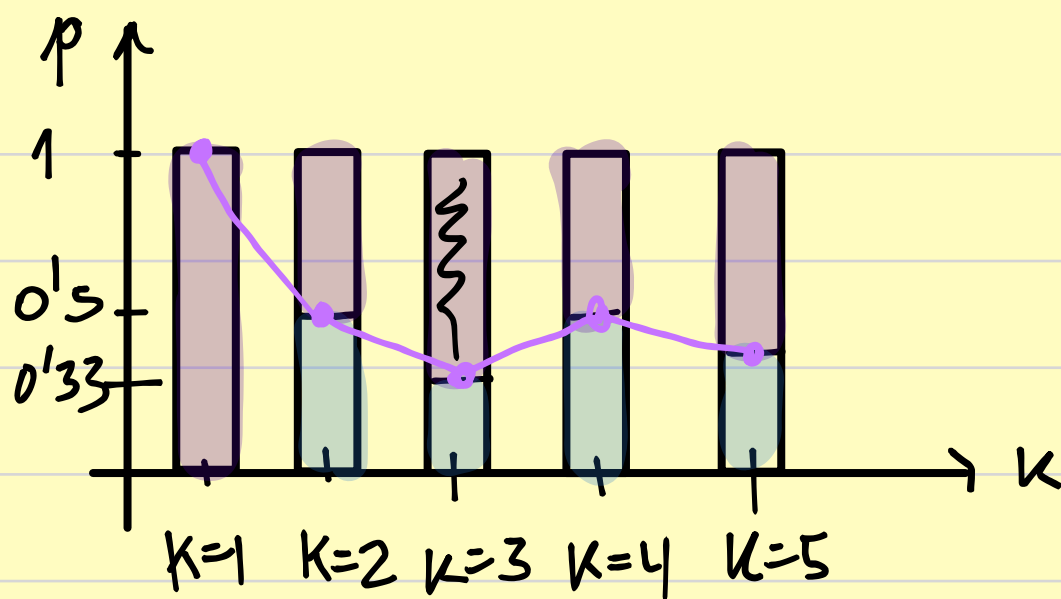
$$k=1: \quad p(X \in N) = \frac{1}{1} = 1 \quad ; \quad p(X \in S) = \frac{0}{1} = 0$$

$$k=2: \quad p(X \in N) = \frac{1}{2} = 0'5 \quad ; \quad p(X \in S) = \frac{1}{2} = 0'5$$

$$k=3: \quad p(X \in N) = \frac{2}{3} = 0'66 \quad ; \quad p(X \in S) = \frac{1}{3} = 0'33$$

$$k=4: \quad p(X \in N) = \frac{2}{4} = 0'5 \quad ; \quad p(X \in S) = \frac{2}{4} = 0'5$$

$$k=5: \quad p(X \in N) = \frac{3}{5} = 0'6 \quad ; \quad p(X \in S) = \frac{2}{5} = 0'4$$



- Das Minimum der Kurve liegt bei $k=3$.
- Somit beschreibt $k=3$ am besten die Trennung der Klassen [Shannon Informationstheorie] und liefert die Entscheidung $\rightarrow X$ gehört zur Klasse **N**.

Übung. (Prüfungssähnlich). OHNE LÖSUNG.

Gegeben sind die Positionen von 6 Werken mit $[x, y]$ Koordinaten. 1) Bitte ermitteln Sie mit K-Means Clustering die 2 relevanten Gruppen um 2 Lager zu positionieren.

Fangen Sie mit Cluster $G_1[w_1, w_2]$ $G_2[w_3, w_4, w_5, w_6]$.

2) Mit Hilfe von KNN entscheiden Sie zur welchen Gruppe Werk x gehört.

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

