Intelligente Systeme
- Soft Computing -

Hochschule für Angewandte Wissenschaften Hamburg
Department Informatik

Dr.-Ing. Sabine Schumann

Roadmap

1. Einleitung
2. Logik: Aussagenlogik, Prädikatenlogik erster Stufe
3. Prolog
4. DCG
5. Problemlösen
6. Constraints
7. Soft Computing:
   Vages, Unsicheres, Unvollständiges Wissen
8. Neuronale Netze
9. Semantische Netze & Frames

Roadmap

Soft Computing

- unsicheres, unscharfes Wissen
- Wahrscheinlichkeiten
- Bedingte Wahrscheinlichkeiten
- Bayessche Formel
- Bayessche Netze
- Fuzzy

Einleitung

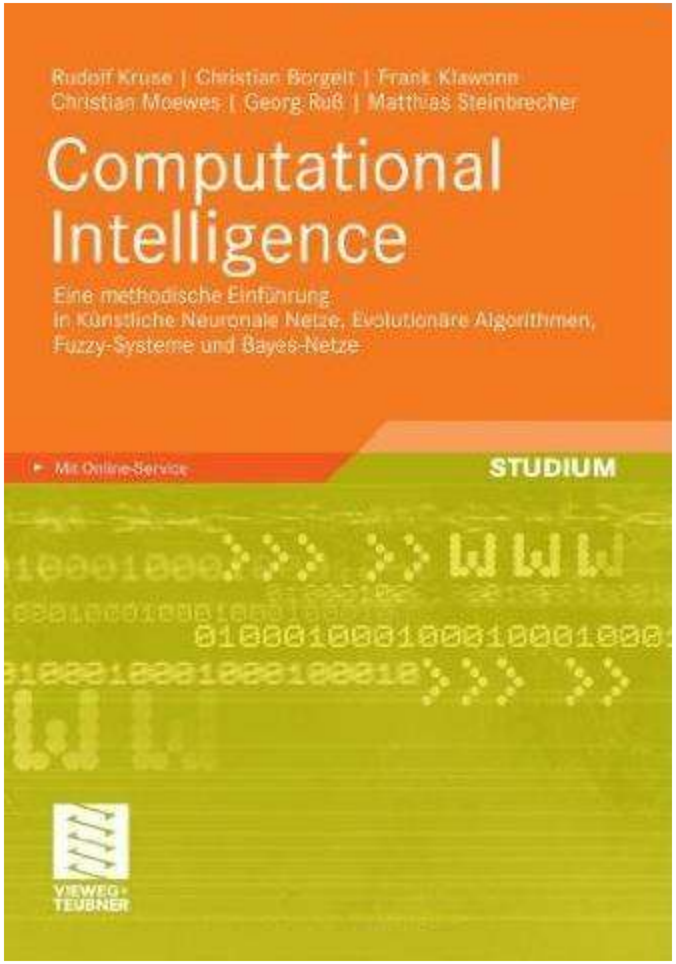
Bücher

Uwe Lämmel, Jürgen Cleve:
***Lehr- und Übungsbuch Künstliche Intelligenz*.**
4. Aufl.,Hanser, München, 2012

Ingo Boersch, Jochen Heinsohn, Rolf Socher:
Wissensverarbeitung: Eine Einführung in die Künstliche Intelligenz für Informatiker und Ingenieure.
2. Aufl., Spektrum Akademischer Verlag, 2007

Einleitung

weiterführende Literatur



Rudolf Kruse (* 1952)

Frank Klawonn (* 1964)

Rudolf Kruse, Christian Borgelt, Frank Klawonn, Christian Moewes, Georg Ruß, Matthias Steinbrecher:
Computational Intelligence: Eine methodische Einführung in Künstliche Neuronale Netze, Evolutionäre Algorithmen, Fuzzy-Systeme und Bayes-Netze.
1. Aufl., Vieweg+Teubner, 2011

Es gibt keine gerade Linie an sich.

Ruler‘s Regel,

Arthur Bloch: Murphy‘s Gesetze in einem Band, 1985

Einleitung

verwendete Quellen

Jürgen Cleve:
Unterlagen zur Veranstaltung KI an der HS Wismar

Kai von Luck:
Unterlagen zur HAW-Veranstaltung „Intelligente Systeme“

Sabine Schumann, Oliver Schumann:
Modellierung von Unschärfe in KONWERK

in Andreas Günter (Hrsg.)
Wissensbasiertes Konfigurieren –
Ergebnisse aus dem Projekt PROKON,
Infix, 1995, S. 97-116

A. Günter (Hrsg.)

Wissensbasiertes
Konfigurieren

Ergebnisse aus dem Projekt PROKON

PROKON

infix

Motivation

- Wieviele Bäume sind ein Wald?
- Wieviele Schrauben sind ein Haufen?
- Eine Blautanne kostet €345, ist das teuer?
- Karl ist 190cm lang. Ist das groß?
- Haye besitzt €300.000. Ist er reich?
- Übermorgen scheint die Sonne.
- Wenn ich die Klausur bestehe, gehe ich bestimmt auf die Party.

Vages Wissen

Häufig vage Aussagen durch

- unscharfe Begriffe bzw.
- unscharfe Aussagen

Mit bisher vorgestellten Methoden nicht modellierbar, denn es ist nicht eindeutig klar, ob solche Aussagen wahr oder falsch sind.

Vages Wissen

Ziel:
auch mit vagen Begriffen Entscheidungen treffen.

Beispiel: Wohnungssuche

- *groß*, *billig*, *stadtnah*, *ruhig*, …

Wohnungsangebote müssen irgendwie bewertet werden … wahr/falsch geht nicht, da die Angebote nicht eindeutig *billig* bzw. *nicht billig* sind.

Bewertung 90% für *billig* zulassen.

Wie verbindet man nun 90% *billig* mit 70% *groß*?

Zudem möchte man Schlussfolgerungen ziehen, ähnlich wie mit der Resolution.

$\forall X\ \text{patient}(X) \land \text{hohesfieber}(X) \Rightarrow \text{maßnahme}\ (X, \text{wadenwickel})$
Ist 38°C hohes Fieber? Nicht zu 100%, eher zu 50%.

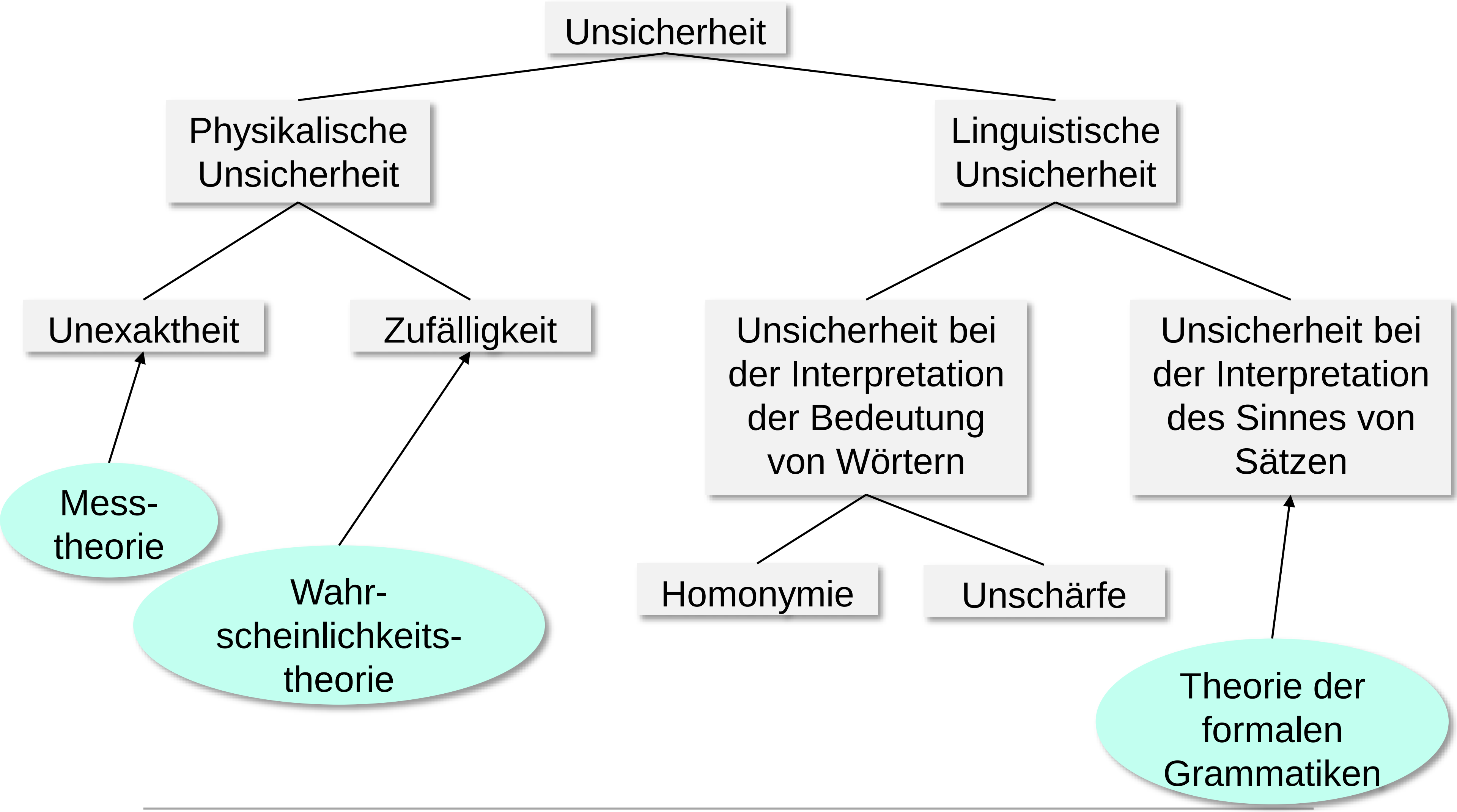
Vages Wissen

Es werden unsichere Aussagen gemacht.
Diese Aussagen können nur wahr oder falsch sein, aber zum jetzigen Zeitpunkt ist keine Aussage darüber möglich, ob wahr oder falsch.

Konsequenzen:

- Zweiwertige Logik ist nicht anwendbar.
- Zulassen von Wahrheitswerten zwischen 0 und 1.
- Üblicher Ansatz:
  Verwendung von Wahrscheinlichkeiten.

Typen von Unsicherheit



Unsicherheit

Physikalische Unsicherheit

Linguistische Unsicherheit

Unexaktheit

Zufälligkeit

Unsicherheit bei der Interpretation der Bedeutung von Wörtern

Unsicherheit bei der Interpretation des Sinnes von Sätzen

Mess- theorie

Wahr- scheinlichkeits- theorie

Homonymie

Unschärfe

Theorie der formalen Grammatiken

Wahrscheinlichkeit

Eine Wahrscheinlichkeit drückt aus, mit welchem Maß man annehmen kann, dass ein Ereignis eintritt.

Beispiel:

Es wird eine gerade Zahl gewürfelt.
… hat eine Wahrscheinlichkeit von 0,5.

Wahrscheinlichkeiten können unterschiedlich interpretiert werden:

- Maß für die Häufigkeit von Ereignissen.
- Subjektives Maß für das Eintreffen von Ereignissen.

Wahrscheinlichkeit

- Wir haben also einen Ereignisraum Ω (frame of discernment).
  in unserem Würfelbeispiel: $\Omega = \{1,2,3,4,5,6\}$
- und Elementarereignisse.
  1, 2, 3, 4, 5, 6
- Jedes Ereignis ist darstellbar als: $A \subseteq \Omega$.
  Ereignis: es wird eine gerade Zahl gewürfelt $A=\{2,4,6\}$
- Eine Aussage heißt wahr, wenn die der Aussage entsprechende Teilmenge das tatsächlich eingetretene Ereignis ω enthält: $\omega \in A$
- Menge aller möglichen Aussagen:
  Menge aller Teilmengen von Ω bezeichnet als 2^{Ω} .
- Ω ist das sichere Ereignis
  (da alle Elementarereignisse in Ω enthalten sind, $A = \Omega$)
- $\varnothing$ ist das unmögliche Ereignis, da die leere Menge kein Elementarereignis beinhaltet.

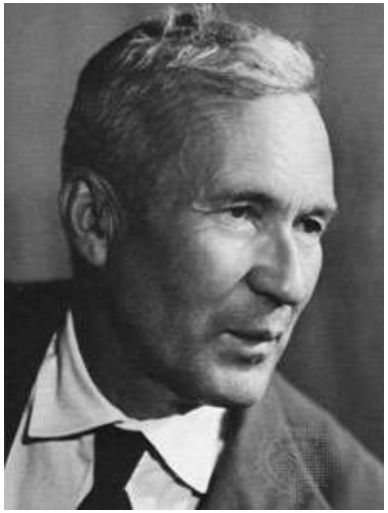
Ereignisse und Logik

Ereignisse können verknüpft werden.
Sind A und B Ereignisse über dem Ereignisraum Ω, so sind es auch:

- $A \cap B$
  Das Ereignis $A \cap B$ tritt genau dann ein,
  wenn A eintritt **UND** auch B eintritt.
  Das entspricht der Aussage $\mathbf{A} \wedge \mathbf{B}$.
- $A \cup B$
  Analog entspricht das Ereignis $A \cup B$ der Aussage $\mathbf{A} \vee \mathbf{B}$.
- A^C (Mengenkomplement)
  entspricht der Negation $\neg\mathbf{A}$.

Wahrscheinlichkeit

Ereignissen sollen Zahlen $\in$ [0,1] zugeordnet werden, die einen **Erwartungswert** für deren Eintreffen darstellen, eine Wahrscheinlichkeit.



Andrei Nikolajewitsch
Kolmogorov
(1903-1987)

Kolmogorov-Axiome:
Sei Ω ein endlicher Ereignisraum.
Eine auf 2^{Ω} definierte Mengenfunktion
$P: 2^{\Omega} \rightarrow [0,1]$ heißt Wahrscheinlichkeit(sfunktion), wenn gilt:

- $P(\Omega)=1$
- $P(A \cup B) = P(A) + P(B)$ für $A,B \subseteq \Omega$ mit $A \cap B = \emptyset$

Wahrscheinlichkeit

Eigenschaften

Für beliebige $A,B \subseteq \Omega$ gilt:

unmögliches Ereignis	$P(\emptyset) = 0$
Monotonie	$A \subseteq B \rightarrow P(A) \leq P(B)$
Subtraktivität	$A \subseteq B \rightarrow P(B \cap A^C) = P(B) - P(A)$
Subadditivität	$P(A \cup B) \leq P(A) + P(B)$
Additivität	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Komplement	$P(A^C) = 1 - P(A)$

Bedingte Wahrscheinlichkeit (1)

Meist sind jedoch abhängige Aussagen vorzufinden:

“Wenn ich die IS Klausur bestehe, dann gehe ich auf die Party.”

“Wenn das Unternehmen X eine positive Jahresbilanz meldet, dann steigt die Unternehmens-Aktie um 5%”

? Was bedeutet ein Wert von 0.8 für diese Aussage?

Bedingte Wahrscheinlichkeit (2)

Betrachten wir
“Wenn das Unternehmen X eine positive Jahresbilanz meldet, dann steigt die Unternehmens-Aktie um 5%”
als logische Implikation:

$$A \Rightarrow B$$

so entspricht das

$$\neg A \vee B$$

Demzufolge bedeutet der Wert 0.8, dass in **80%** der Fälle das Unternehmen X keine positive Jahresbilanz meldet oder die Aktie um 5% steigt.

➡ D.h. die 0.8 gelten für die gesamte Aussage.

Bedingte Wahrscheinlichkeit

Formulieren wir die Aussagen nur leicht verändert, aber mit erheblichem semantischen Unterschied:

“Wenn ich die IS Klausur bestehe, dann gehe ich sicherlich (0.75) auf die Party.”

“Wenn eine positive Jahresbilanz gemeldet wird, dann steigt die Aktie mit einer Wahrscheinlichkeit von 0.8 um 5%”

Solche Aussagen, im einfachsten Fall der Form

- if A then B with p oder auch
- $A \overset{p}{\Rightarrow} B$,

nennt man bedingte Wahrscheinlichkeit.

Bedingte Wahrscheinlichkeit

Interpretation

Warum ist der Unterschied so erheblich?

“Wenn eine positive Jahresbilanz gemeldet wird, dann steigt die Aktie mit einer Wahrscheinlichkeit von 0.8 um 5%”

Betrachten wir folgenden Fall:

- 7 mal keine positive Bilanz, 3 mal positive Bilanz
- Aktie steigt in einem dieser 3 Fälle um 5%

- Erste Interpretation ($\neg A \vee B$):
  0.8 erfüllt (logische Implikation in 8 von 10 Fällen)
- Zweite Interpretation (bedingte Wahrscheinlichkeit $A \overset{p}{\Rightarrow} B$):
  0.8 nicht erfüllt (Aktie stieg in nur 1 von 3 Fällen)

Bedingte Wahrscheinlichkeit

Sei B ein Ereignis mit P(B) > 0. Dann heißt

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

bedingte Wahrscheinlichkeit von A unter der Bedingung B.

Falls P(B) = 0, ist die bedingte Wahrscheinlichkeit nicht definiert.

Bedingte Wahrscheinlichkeit

Beispiel:
Welchen Wahrscheinlichkeitswert erhält folgende Aussage:

“Wird eine gerade Zahl mit einem nicht manipulierten Würfel gewürfelt, dann ist es eine 6.”

Wahrscheinlichkeit

- für das Ereignis GZ = {2, 4, 6}:
  $P(GZ) = \frac{1}{2}$,
- für das Ereignis Z = {6}:
  $P(Z) = \frac{1}{6}$.

Die bedingte Wahrscheinlichkeit:

$$P(Z|GZ) = \frac{P(Z \cap GZ)}{P(GZ)} =$$

Bedingte Wahrscheinlichkeit

Allgemeine Multiplikationsregel

Der Würfel … wenn eine gerade Zahl gewürfelt wurde, ist nun bekannt

$P(\{i\}) = P(\{i\}|GZ)$

$P(\{1\}) = P(\{3\}) = P(\{5\}) = 0$

$P(\{2\}) = P(\{4\}) = P(\{6\}) = \frac{1}{3}$

Umstellen der Formel für bedingte Wahrscheinlichkeiten

$P(A|B) = \frac{P(A \cap B)}{P(B)}$: $P(A \cap B) = P(A|B) * P(B)$

Diese Formel läßt sich nun direkt erweitern, erstmal auf 3 Mengen:

$$P(A_3 \cap A_2 \cap A_1) = P(A_3|A_2 \cap A_1) * P(A_2 \cap A_1)$$
$$= P(A_3|A_2 \cap A_1) * P(A_2|A_1) * P(A_1)$$

Generalisierung auf n Mengen:

allgemeine Multiplikationsregel (auch Verkettungsregel)

Gegeben sind die Ereignisse $A_1, \ldots, A_n \subseteq \Omega$ mit $P(A_1 \cap \ldots \cap A_{n-1}) > 0$, dann gilt:

$P(A_1 \cap \ldots \cap A_n) =$

$P(A_n|A_1 \cap \ldots \cap A_{n-1}) * \ldots * P(A_3|A_1 \cap A_2) * P(A_2|A_1) * P(A_1)$

oder anders: $\prod_{i=1}^{n} P(A_i \,|A_{i-1} \cap \ldots \cap A_1)$

Deduktion und Induktion

Bei der Deduktion wird aus einem eingetretenen Ereignis auf ein davon abhängiges Ereignis geschlossen.

Bei der Induktion wird aus einem eingetretenen Ereignis auf ein erzeugendes Ereignis zurück geschlossen.

Bayessche Formel

Erneut:
Umstellen der Formel für bedingte Wahrscheinlichkeiten

$P(A|B) = \frac{P(A \cap B)}{P(B)} : \quad P(A \cap B) = P(A|B) * P(B)$

Betrachten wir nun die bedingte Wahrscheinlichkeit von B unter der Bedingung A, erhalten wir:

$P(B \cap A) = P(B|A) * P(A)$

Der Mengendurchschnitt ist kommutativ …

$P(B) * P(A|B) = P(A) * P(B|A)$

Thomas Bayes
170?-1761
Quelle: wikipedia

… und wir erhalten die Bayessche Formel:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Mit der Bayesschen Formel kann von einem eingetretenen Ereignis auf ein erzeugendes Ereignis geschlossen werden (Induktion).

Bayessche Formel

Akku-Beispiel I (1)

aus Boersch, Heinsohn, Socher: Wissensverarbeitung, Spektrum Verlag

- akkubetriebener autonomer Roboter, dessen Fehlerrate vom Ladezustand des Akkus abhängt.
- schwacher Akku – Fehlerrate nimmt drastisch zu
- Erfahrungen der letzten Wochen:
  - Der Roboter macht bei ca. 20% seiner Aktionen Fehler.
    → P(Fehler) = 0.2
  - Bei schwachem Akku steigt die Fehlerrate auf 80%.
    → P(Fehler | schwach) = P(Fehler | nicht voll) = 0.8
  - Im Durchschnitt ist der Akku in einer von 10 Minuten schwach.
    → P(schwach) = P(nicht voll) = 0.1

? Wie hoch ist die Wahrscheinlichkeit, dass der Akku schwach ist, wenn bei einer Roboteraktion ein Fehler gemacht wird?

→ P(schwach | Fehler)

Bayessche Formel

Akku-Beispiel I (2)

Bayessche Formel:

$$P(A/B) = \frac{P(B|A) * P(A)}{P(B)}$$

$$P(schwach/Fehler) = \frac{P(Fehler|schwach) * P(schwach)}{P(Fehler)}$$

$$P(schwach/Fehler) = \frac{0.8 * 0.1}{0.2} = \mathbf{0.4}$$

gegeben:
P(Fehler) = 0.2
P(schwach) = 0.1
P(Fehler | schwach) = 0.8

gesucht:
P(schwach | Fehler)

Bayessche Formel

Akku-Beispiel II (1)

Aufgabenstellung leicht abgeändert!!!

- akkubetriebener autonomer Roboter, dessen Fehlerrate vom Ladezustand des Akkus abhängt.
- schwacher Akku – Fehlerrate nimmt drastisch zu
- Erfahrungen der letzten Wochen:
  - Der Roboter macht mit **vollem** Akku bei ca. 20% seiner Aktionen Fehler.

    ➡ **P(Fehler | voll) = 0.2**
  - Bei schwachem Akku steigt die Fehlerrate auf 80%.

    ➡ P(Fehler | schwach) = P(Fehler | nicht voll) = 0.8
  - Im Durchschnitt ist der Akku in einer von 10 Minuten schwach.

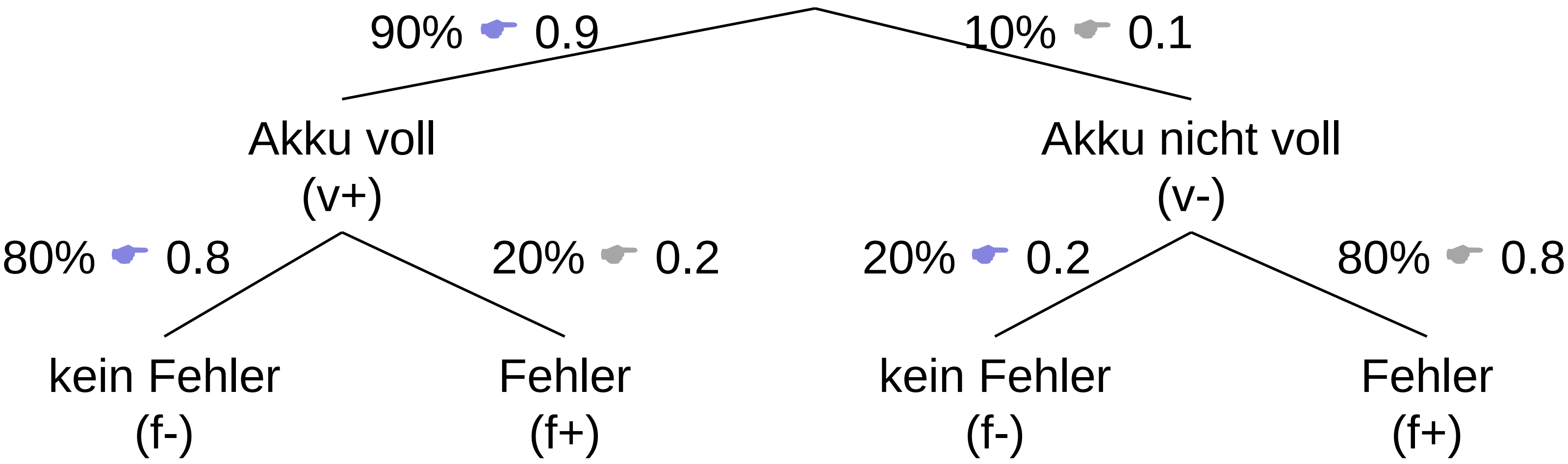
    ➡ P(schwach) = P(nicht voll) = 0.1

? Wie hoch ist die Wahrscheinlichkeit, dass der Akku schwach ist, wenn bei einer Roboteraktion ein Fehler gemacht wird?

➘ P(schwach | Fehler)

Bayessche Formel

Akku-Beispiel II (2)



90% ☛ 0.9

10% ☛ 0.1

Akku voll
(v+)

Akku nicht voll
(v-)

80% ☛ 0.8

20% ☛ 0.2

20% ☛ 0.2

80% ☛ 0.8

kein Fehler
(f-)

Fehler
(f+)

kein Fehler
(f-)

Fehler
(f+)

$P(v+ \cap f-)=\mathbf{0.72}$ $P(v+ \cap f+)=\mathbf{0.18}$ $P(v- \cap f-)=\mathbf{0.02}$ $P(v- \cap f-)=\mathbf{0.08}$

anders:

	v+	v-
f+		
f-		

Bayessche Formel

Akku-Beispiel II (3)

gegeben:
P(Fehler | voll) = 0.2
P(schwach) = 0.1
P(Fehler | schwach) = 0.8

errechnet:
P(Fehler) = 0.26

gesucht:
P(schwach | Fehler)

? Wie hoch ist die Wahrscheinlichkeit, dass der Akku schwach ist, wenn bei einer Roboteraktion ein Fehler gemacht wird?

Bayessche Formel:

$$P(A/B) = \frac{P(B|A)*P(A)}{P(B)}$$

$$P(schwach/Fehler) = \frac{P(Fehler|schwach)*P(schwach)}{P(Fehler)}$$

$$P(schwach/Fehler) = \frac{0.8*0.1}{0.26} = \mathbf{0.31}$$

Bayessche Formel

und noch ein Beispiel(1)

Wir betrachten ein Heimspiel von St. Pauli.

- Wenn Pauli gewinnt oder unentschieden spielt, gibt es die nächste Eintrittskarte **günstiger**.
- **Wahrscheinlichkeit** für den Ausgang eines Heimspiels von Pauli:
  - Sieg: 0.85
  - Unentschieden: 0.1
  - Niederlage: 0.05

Wie hoch ist die Wahrscheinlichkeit, dass Pauli gewonnen hat, wenn wir für das nächste Heimspiel die Karte günstiger bekommen?

Wir suchen: P(Sieg | Karte günstiger)

P(Karte günstiger) = 0.85 + 0.1 = 0.95

P(Sieg) = 0.85

P (Karte günstiger | Sieg) = 1

$$P(Sieg \mid Karte\ günstiger) = \frac{P(Karte\ günstiger\ |Sieg) * P(Sieg)}{P(Karte\ günstiger)}$$

*P(Sieg | Karte günstiger) = 1 * 0.85 / 0.95 = 0.8947* $\approx$ **0.89**

Produkträume (1)

Zurück zur Party:

“Wenn ich die IS Klausur (k) bestehe, dann gehe ich sicherlich (0.75) auf die Party (p).
Sollte ich die Klausur (k) nicht bestehen, kann es aber auch sein (0.5), dass ich auf die Party (p) gehe.”

Ω_{party} = {p+, p-}
$\Omega_{klausur}$ = {k+, k-}
Ω = $\Omega_p \times \Omega_k$
= {(p+, k+), (p-, k+), (p+, k-), (p-, k-)}

k+ $\overset{0.75}{\Longrightarrow}$ p+, k- $\overset{0.5}{\Rightarrow}$ p+

$P(\omega_p \mid \omega_k)$	p+	p-
k+	3/4	1/4
k-	1/2	1/2

Verkürzte Schreibweise:
$P(\omega_p \mid \omega_k)$ für $P(\{\omega_p\} \mid \{\omega_k\})$
$P(\omega_p, \omega_k)$ für $P(\{(\omega_p, \omega_k)\})$

Da $P(k+) = P(k-) = \frac{1}{2}$
lassen sich mit Hilfe von $P(\omega_p, \omega_k) = P(\omega_p \mid \omega_k) * P(\omega_k)$
die Wahrscheinlichkeiten berechnen:

Produkträume (2)

$P(\omega_p, \omega_k)$	p+	p-
k+	3/8	1/8
k-	1/4	1/4

Auch für Produkträume gilt:

Die Summe der Wahrscheinlichkeiten der Elementarereignisse muss 1 ergeben:

$$P(\Omega) = \sum_{\omega \in \Omega} P(\omega) = \sum_{(\omega_k, \omega_p) \in \Omega_k \times \Omega_p} P(\omega_p,\ \omega_k) = 1$$

Nun können wir Randverteilungen berechnen:

z. Bsp.: Mit welcher Wahrscheinlichkeit gehe ich überhaupt zur Party?

$$P(\omega_p) \quad = \sum_{\omega_k \in \Omega_k} P(\omega_k,\ \omega_p)$$

$$P(p+) = P(k+, p+) + P(k-, p+) \ = \frac{1}{4} + \frac{3}{8} = \frac{5}{8}$$

$$P(p-) \ = P(k+, p-) + P(k-, p-) \quad = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}$$

Unabhängigkeit von Ereignissen

Notationen:

- $A_1, \ldots, A_n$ Variablen bzw. Ereignisse über
- $\Omega_{A_1}, \ldots, \Omega_{A_n}$ Wertebereich von A_i
- $a_1, \ldots, a_n$ Elemente

Variablen können Elemente oder auch Mengen von Elementen des Wertebereiches denotieren, bei wenigen Variablen wird auf die Indizierung verzichtet.

Wenn für 2 Variablen A und B gilt: $P(A|B)=P(A)$, dann hat B anscheinend keinen Einfluss auf die Wahrscheinlichkeit von A, man bezeichnet A und B als unabhängig.

Übertragen auf die Produkträume:
Zwei Ereignisse A und B heißen statisch unabhängig, wenn
$P(A,B) = P(A|B) * P(B) = P(A) * P(B)$

Beispiel Würfel:
Im ersten Versuch wird eine gerade Zahl gewürfelt (und) im zweiten Versuch wird eine ungerade Zahl gewürfelt. P(A,B) = 0.5*0.5 = 0.25

Bedingte Unabhängigkeit von Ereignissen

Zwei Ereignisse A und B heißen bedingt unabhängig unter der Bedingung C, wenn

$P(A,B|C) = P(A|C) * P(B|C)$

es gilt auch:

$P(A|B,C) = P(A|C)$

Beispiel:

Fahrzeuge:
Typ (T) bestimmt Gewicht (G) und Benzinverbrauch (B).
Gewicht und Benzinverbrauch sind voneinander abhängig
$P(G|B) \neq P(G) * P(B)$
Wenn nun der konkrete Fahrzeugtyp bekannt ist, gilt obige Gleichung, d.h. bei bekanntem Fahrzeugtyp sind Gewicht und Benzinverbrauch unabhängig.
Interessiert mich das Gewicht des Fahrzeuges, benötige ich nur den Fahrzeugtyp, jedoch nicht den Benzinverbrauch ($P(G|B,T) = P(G|T)$)

Bayessche Formel

Es kann bei der Anwendung der Baysschen Formel zu diversen Paradoxen kommen.

- häufig sind die wahrscheinlichkeitstheoretischen Voraussetzungen nicht erfüllt bzw. prüfbar widersprüchliche subjektive Werte können zu P(A|B) >1 führen
- unabhängige vs. abhängige Ereignisse:
  - unabhängig:
    Im ersten Versuch wird eine gerade Zahl gewürfelt (und) im zweiten Versuch wird eine ungerade Zahl gewürfelt.
    P=0.5*0.5=0.25
  - abhängig:
    Im ersten Versuch wird eine gerade Zahl gewürfelt (und) im ersten Versuch wird eine ungerade Zahl gewürfelt.
    P=0 (beide Ereinisse können nicht gleichzeitig auftreten)
- weitere:
  http://www2.hs-fulda.de/~grams/dnkfln.htm#_Bayes-Schätzung

Bayessche Formel

Anwendungen

- Statistik,
- Hypothesenwissen modellieren,
- Schlussfolgerungen (Schließen) aus unsicherem Wissen,
- Qualitätsmanagement,
- Entscheidungstheorie,
- Verkehrsverteilung-Vorhersage,
- …

Bayessche Netze

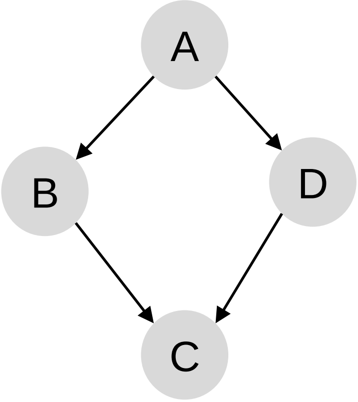
- ab Mitte der 70er Jahre Kritik, da Unsicherheitsmaße und Inferenzverfahren häufig rein intuitiv begründet waren (insbesondere in der Medizin).
- Einbeziehung graphischer Faktoren
- Klassen von Wahrscheinlichkeiten
- Der neue Ansatz wird als
    - Bayessche Netze (*Bayesian networks*) bzw.
    - zerlegbare graphenbasierte Modelle (*decomposable graphic models*) bzw.
    - Belief-Netze (belief networks)

  bezeichnet.

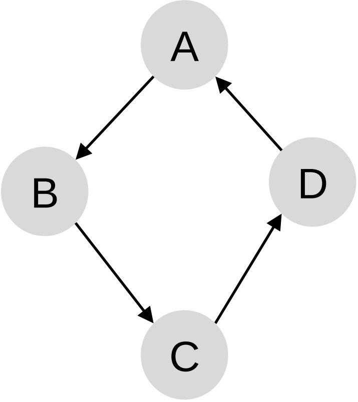
Ein Bayessches Netz ist ein gerichteter azyklischer Graph (DAG), dessen Knoten Variablen und dessen Kanten die Existenz direkter Abhängigkeiten zwischen Knoten denotieren. Die Stärke solcher Abhängigkeiten wird auf Basis **bedingter Wahrscheinlichkeiten** angegeben.

Bayessche Netze

- Knoten des Netzes: Zufallsvariablen.
- Intuitive Bedeutung einer Kante von Knoten A zu Knoten B:
  A hat **direkten Einfluss** auf B.
- Jeder Knoten hat eine Tabelle mit bedingten Wahrscheinlichkeiten, die den Einfluss der Eltern (Vorgängerknoten) auf den Knoten darstellt.
- Graph ist zyklenfrei: gerichteter azyklischer Graph (DAG).
- Diskrete Zufallsvariablen mit einer endlichen Anzahl von Zuständen.



erlaubt



nicht erlaubt

Bayessche Netze

Beispiel(1)

Herr White ist nicht zu Hause. Sein Nachbar Herr Black ruft ihn an und teilt ihm mit, dass die Alarmanlage in Herrn Whites Garage ausgelöst wurde. Die Nachbarin Frau Pink ruft ebenfalls an und teilt Herrn White das gleiche mit.
Die Alarmanlage wird manchmal durch Gewitter ausgelöst, aber natürlich auch durch einen Einbrecher.

Variablen:
Alarm, Anruf Frau Pink, Anruf Herr Black, Einbrecher, Gewitter

Gewitter und Einbrecher sind voneinander unabhängig:
P(Gewitter|Einbrecher) = P(Gewitter)
P(Einbrecher|Gewitter) = P(Einbrecher)

Frau Pink und Herr Black rufen an, wenn sie den Alarm gehört haben.

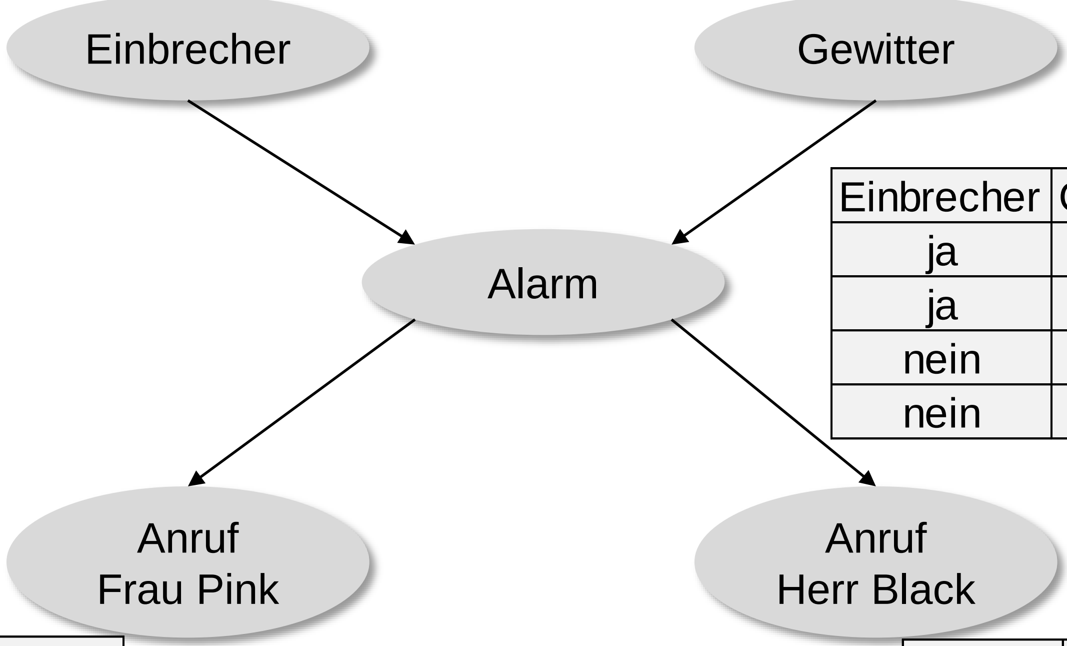
Bayessche Netze

Beispiel(2)

Dafür bauen wir nun eine Graphik mit dazugehörigen Wahrscheinlichkeiten (aus Beobachtungen der Realität):

P(Einbrecher) = 0.001

P(Gewitter) = 0.002



Einbrecher

Gewitter

Alarm

Anruf
Frau Pink

Anruf
Herr Black

Einbrecher	Gewitter	P(Alarm)
ja	ja	0.95
ja	nein	0.94
nein	ja	0.29
nein	nein	0.001

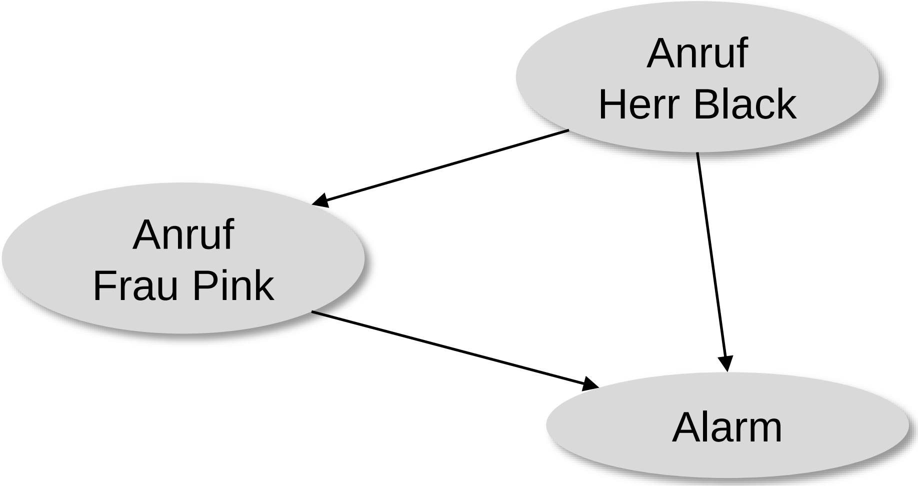
Alarm	P(Pink)
ja	0.7
nein	0.01

Alarm	P(Black)
ja	0.9
nein	0.05

Bayessche Netze

Die Kompaktheit eines Netzes hängt sehr stark von der Reihenfolge der eingefügten Variablen ab, dabei muss die Abhängigkeit beachtet werden!!!

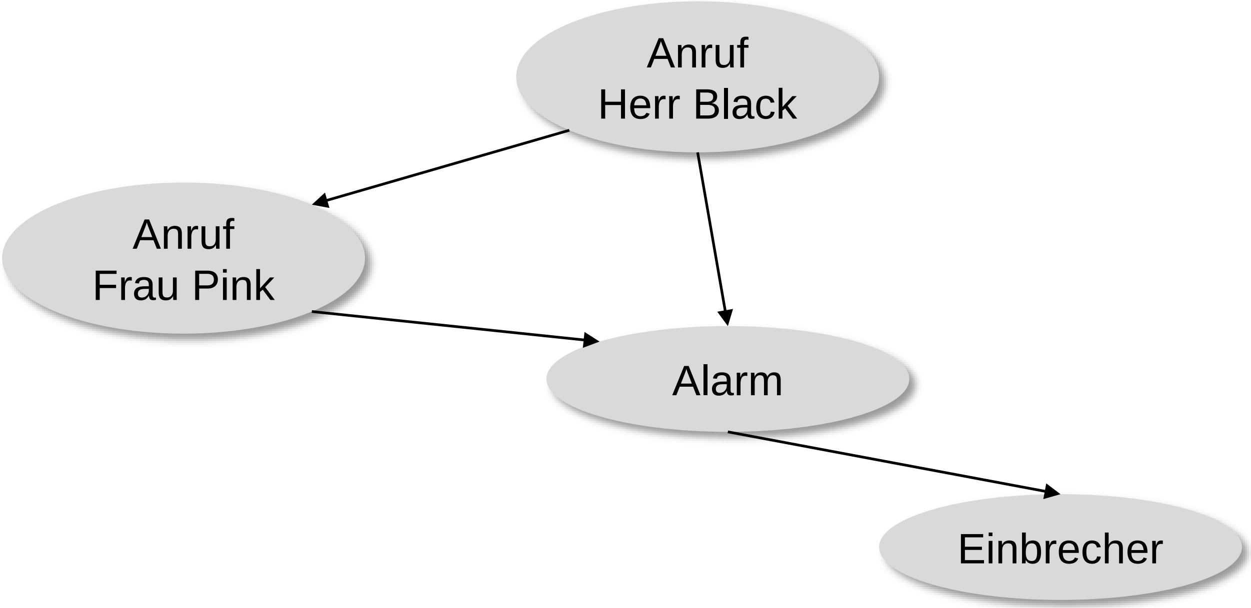
Beispiel für ungünstige Reihenfolge:
Wählt man den *Anruf Herr Black* als erste und den *Anruf Frau Pink* als zweite Variable aus, muss man überprüfen, ob Anruf Frau Pink vom Anruf Herr Black **unabhängig** ist.
P(Anruf Pink | Anruf Black) = P(Anruf Pink)
NEIN
Hinzufügen von Alarm



Bayessche Netze

Beispiel(4)

Hinzufügen von *Einbrecher*:
Prüfen, ob *Einbrecher* von *Anruf Frau Pink, Anruf Herr Black und Alarm* **bedingt abhängig** ist.
Da *Einbrecher* unabhängig von den Anrufen ist:
P(Einbrecher | Anruf Pink, Anruf Black, Alarm) =
P(Einbrecher | Alarm)



Bayessche Netze

Beispiel(5)

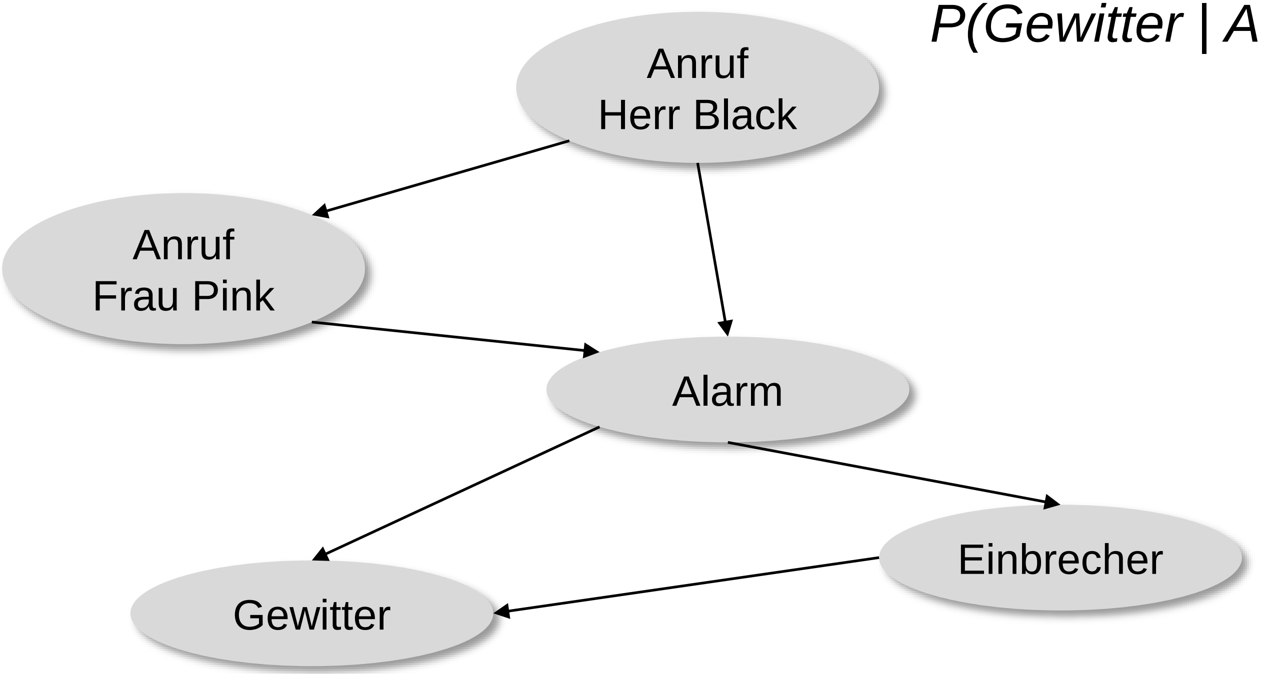
Hinzufügen von *Gewitter*:

Prüfen, wovon *Gewitter* **bedingt abhängig** ist.

Da bei einem Alarm die Information über einen Einbrecher relevant ist, hängt *Gewitter* sowohl von *Alarm* als auch von *Einbrecher* ab.

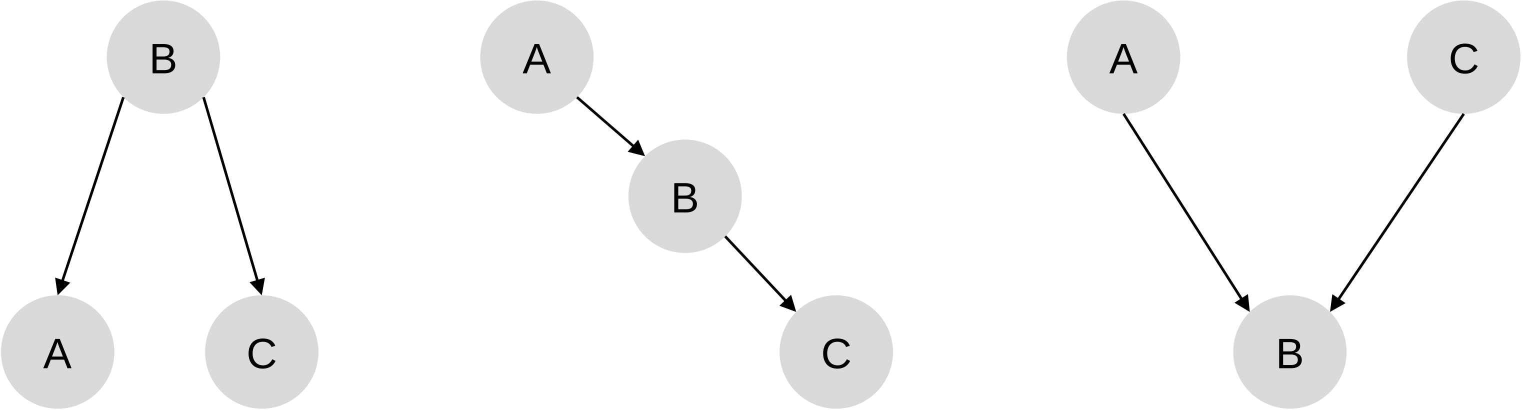
P(Gewitter | Anruf Pink, Anruf Black, Alarm, Einbrecher) =

P(Gewitter | Alarm, Einbrecher)



Abhängigkeitsstrukturen(1)

Wichtig scheinen also die Un- und Abhängigkeiten zu sein.
Wir unterscheiden (von B ausgehend):



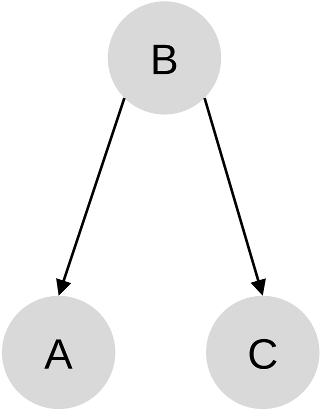
divergierende Kanten:
P(A,C|B) = P(A|B) * P(C|B)
P(C|A,B) = P(C|B)

kaskadierende Kanten:

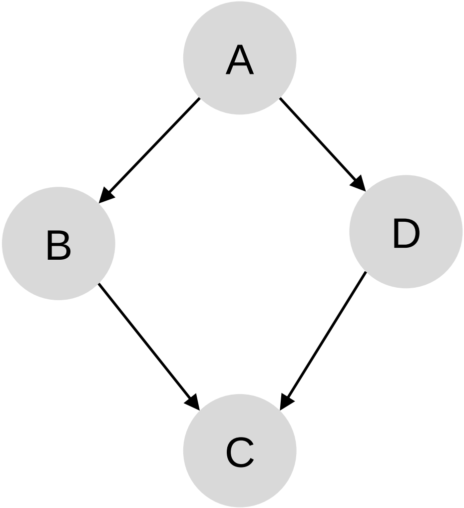
konvergierende Kanten:
P(A|C) = P(A)
P(A|B,C) ≠ P(A|B),
d.h. nur unabhängig, wenn
B unbekannt

Abhängigkeitsstrukturen(2)

Beispiel:
Sind A, C bei gegebenem B bedingt unabhängig?
P(A,C|B)=P(A|B)*P(C|B)



Ja



Nein, weil Weg über D, d.h. A, C sind bedingt unabhängig, wenn B und D gegeben sind.

Was wird nun durch ein Bayessches Netz besser?
Bisher Multiplikationsregel:

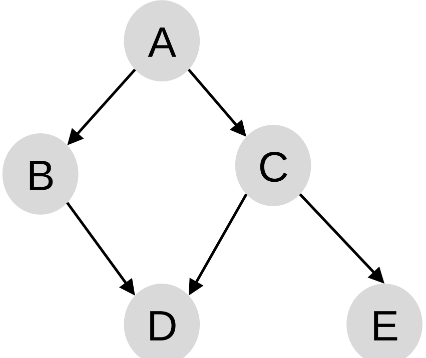
$$\prod_{i=1}^{n} P(A_i \,|A_{i\,-\,1} \,\cap\, \dots \,\cap A_1)$$

- n Variablen, bei zweiwertigen Variablen (stark oder schwach, ja oder nein, …) Ausrechnen von 2^n Wahrscheinlichkeiten für jede Belegung mit Hilfe der Multiplikationsformel.
- Aufgrund der statischen und bedingten Unabhängigkeiten kann die Produktkette verkürzt werden.
- Von einem Knoten V_i ausgehend, interessieren nur noch die Vorgängerknoten Eltern(V_i).
- Die Formel vereinfacht sich:

$$\prod_{i=1}^{n} P(V_i \,|Eltern(Vi))$$

Bayessche Netze

Berechnen Beispiel



Beispiel:

- 5 Variablen A, B, C, D, E
- allgemeine Multiplikationsformel für eine konkrete Belegung a, b, c, d, e:
  P(a,b,c,d,e) = P(e|a,b,c,d) * P(d|a,b,c) * P(c|a,b) * P(b|a) * P(a)
- Dank der modellierten Abhängigkeiten im Bayesschen Netz reduziert sich das auf:
  P(a,b,c,d,e) = P(e|c) * P(d|b,c) * P(c|a) * P(b|a) * P(a)

Damit wird nun für alle verschiedenen Variablenbelegungen im Produktraum die Wahrscheinlichkeit ausgerechnet.

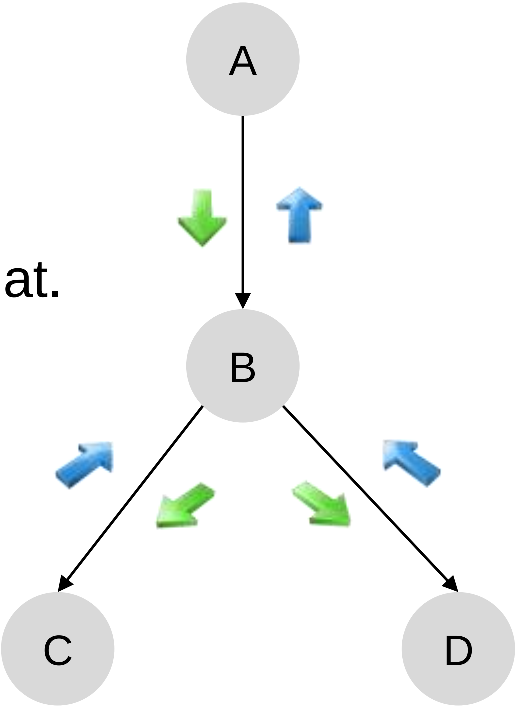
Sobald eine Variable auf einen Wert festgelegt wird, geht die Rechnerei erneut los (bspw. E: P(e+=1), P(e-=0)), da sich dieses Zusatzwissen auf andere Variablen im Netz auswirkt.

Propagierung

Propagierung (Ausbreitung):
Beschleunigung der Wahrscheinlichkeitsberechnung für alle Knoten im Bayesschen Netz.

- betrachtet werden die Einflussfaktoren einer Variable V auf ihre Eltern und
- betrachtet werden die Einflussfaktoren einer Variable V auf ihre Kinder.
- Damit lassen sich die neuen Wahrscheinlichkeitswerte für die Eltern/Kinder berechnen,
- diese geben ihre Änderung dann entlang (noch nicht abgearbeiteter Kanten !!! (**d.h. es wird nicht zurück propagiert!**)) weiter an ihre Eltern/Kinder, bis das ganze Netz durchpropagiert ist.

Propagierung

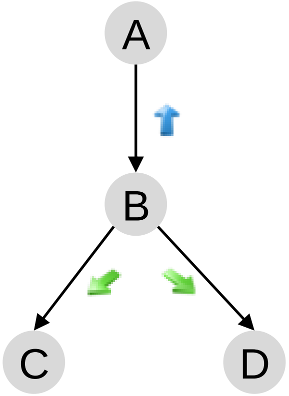


Datenfluss:

Jedem Knoten ist bekannt, welchen Einfluss er auf Elternknoten ➡ bzw. auf seine Kinder ➡ hat.

Dieser Einfluss wird durch die bedingten Wahrscheinlichkeiten bestimmt und kann unterschiedlich groß sein
(z.B. das Vorliegen von Husten beeinflusst die Wahrscheinlichkeiten für Lungenkrebs, Tbc, Asthma, Lungenentzündung, Grippe, “hat sich verschluckt” in unterschiedlichem Maße).

Beispiel Datenfluss, wenn Wert für B bekannt wird:



Zusammenfassung Bayessche Netze

- Bayessche Netze sind eine Kombination von Graphentheorie und Wahrscheinlichkeitstheorie.
- Wissen wird als Ansammlung von Wahrscheinlichkeiten ausgedrückt.
- Bei Polytree exakte Inferenz in linearer Zeit möglich, ansonsten exponentielle Laufzeit.
- Anwendungen:
  Entscheidungsfindung,
  Wissensrepräsentation,
  Wettervorhersagen,
  Diagnosen, …

Fuzzy …

Fuzzy …
englisch für verwischt, verschwommen, unbestimmt.

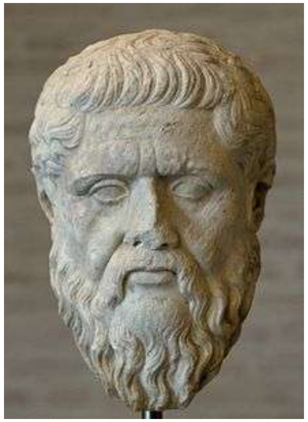
Fuzzy Logik:

- Modellierung von Unsicherheiten und Unschärfe.
- Verallgemeinerung der zweiwertigen BOOLEschen Logik.
- Basiert auf
    - Fuzzy-Mengen (fuzzy sets),
    - Zugehörigkeitsfunktionen (fuzzy functions) und
    - logische Operationen.
- Weiterhin:
  Methoden zur Fuzzyfizierung bzw. Defuzzyfizierung.

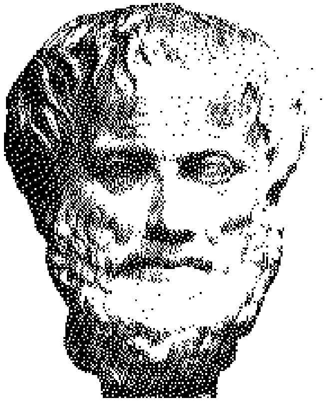
Historie I

Die Wurzeln reichen zurück in die griechische Philosophie.
Platon stellte Überlegungen dazu an, dass zwischen den Begriffen *wahr* und *falsch* ein dritter liegen müsse, während sein Zeitgenosse **Aristoteles** darauf beharrte, dass eine Aussage wahr oder falsch sei, damit die Präzision der Mathematik gegeben sei.

(Das Gesetz vom ausgeschlossenen Dritten … bestimmte die Entwicklung mathematischer und logischer Systeme in den nächsten zwei Jahrtausenden …)



Platon (427 – 347 v. Chr.)
Bronzebüste in der Glyptothek München



Aristoteles (384 – 322 v. Chr.)

Historie II

Moderne Philosophen wie Georg **Hegel** und Bertrand **Russel** nahmen Platons Vermutung wieder auf.

Georg Hegel
(1770-1831)

Russel:

„The law of the excluded middle is true, when precise symbols are employed, but it is not true, when symbols are vague, as, in fact, all symbols are.“

Bertrand Russell
(1872-1970)

Historie III

Jan Łukasiewicz

- systematische Alternative zur zweiwertigen Logik
- „wahr (1)“, „possible (1/2)“, „falsch(0)“
- später vier und fünfwertige Logik
- schließlich unendlichwertige Logik „alle Zahlen im Intervall [0,1]“

Jan Łukasiewicz
(1878-1956)

Historie IV

Max Black

- Verfahren zur numerischen Darstellung von Unschärfe von Symbolen (in Anlehnung an Russel)
- Definition der Ungenauigkeit oder Vagheit eines Symbols unter Zuhilfenahme des Komplements:
  es gibt mindestens ein Element, das weder vollständig zum Symbol noch vollständig zum Komplement gehört
- Menge der Elemente, die nicht eindeutig zugeordnet werden können:
  „frings“ = Fransen

Max Black
(1909-1988)

Historie V

Lotfali Asker Zadeh (*1921)

Lotfi A. Zadeh gilt als Begründer der Fuzzy-Mengen und Fuzzy Logik.

- lehrte ab 1959 an der Universität von Kalifornien (Berkeley)
- stellte 1965 erstmals in einem Artikel das Konzept der Fuzzy Logik vor („Fuzzy-Sets“).
- verbindet darin Blacks Idee der „frings“ mit Łukasiewiczs unendlichwertiger Logik

Theorie der unscharfen Mengen wurde durch die Beobachtung von Lotfi Zadeh ausgelöst, dass Menschen anscheinend in Kategorien denken und kommunizieren, die sich von den in Mengenlehre und Logik verwendeten (dualen) Strukturen unterscheiden.

Dies war zwar schon früher erkannt worden, aber Zadeh war der erste, den diese Beobachtung zur Formulierung einer neuen Theorie veranlasste.

Abgrenzung

Die Fuzzy Logik ist keine mehrwertige Logik, auch wenn sie im engeren Sinne als solche gedeutet werden kann (statt nur Wahrheitswerte von 0 und 1 nun reelle Zahlen im Intervall [0,1]).

Zadeh ging jedoch nicht vom Begriff der Logik aus, sondern betrachtete Fuzzy-Mengen.
Er wollte die Zugehörigkeit von Objekten zu Mengen graduell über numerische Werte zwischen 0 und 1 angeben. Er schuf damit eine Formalisierung von unbestimmten Begriffsumfängen
(z.Bsp. Begriff “groß”).

Erst sein Doktorand Joseph Goguen erwähnte erstmals “Logik der Unschärfe”.

Joseph Goguen
(1941 – 2006)

Foto: http://cseweb.ucsd.edu/~goguen/

Historie VI

80-er Jahre

Die Idee wurde erst in den 80er Jahren von japanischen Wissenschaftlern aufgegriffen, als man die Bedeutung der Idee für die Regelung von Systemen erkannte. Es kam zur sogenannten *Fuzzy-Welle*. 1990 wird in Japan auch das **Fuzzy-Logik-Jahr** genannt.

- erste fuzzy-geregelte Waschmaschinen, Staubsauger
  → höherer Bedienkomfort, Aufsehen als „denkende" Konsumgüter
- erster Fuzzy-Regler in großtechnischer Anwendung
  → Vermutung als neues Universalwerkzeug der Regelungstechnik
- Skepsis in Europa, besonders in Deutschland
  → Qualitätsverbesserung bei Fuzzy-Einsatz oft durch zusätzliche Sensorik erreicht

Historie VII

90-er Jahre

In Europa begann die Entwicklung erst in den 90ern.

- Fuzzy-Logik und Fuzzy-Control von Hochschulen aufgegriffen
  → Fuzzy-Technologie keine vorübergehende Mode, sondern Revolution, die mit Denkgewohnheiten bricht und praktisch vorzeigbare Vorteile bringt
  → neuer Zweig der Regelungstechnik ist entstanden
  → Querschnittswissenschaft, die auch in nichttechnischen Bereichen Vorteile verspricht

Ziele

die man sich durch Fuzzy verspricht:

- Modellverbesserung (Relaxierung),
- Komplexitätsreduktion,
- Modellierung von Unsicherheit,
- bedeutungserhaltendes Schließen,
- …

wichtige Anwendungen – methodische Sicht

Mathematik,
Ökonomie,
Psychologie,
…

Netzwerkplantechnik,
Optimierung,
Clustern,
mehrkriterielle Analyse,
Datenanalyse,
…

Unscharfe Datenbanken,
Fuzzy-Programmiersprachen,
…

Fuzzy Control,
Unscharfe Datenanalyse,
Fuzzy-Expertensysteme,
…

Anwendungsgebiete

- Wirtschaftswissenschaften
- Entscheidungstheorie
- Finanzdienstleistungen
- Mathematik
  - Topologie,
  - Algebra,
  - Graphen & Netze,
  - Optimierung, …
- Naturwissenschaften
  - Physik (Quantenmechanik, Strömungsdynamik),
  - Chemie (Wirkstrukturanalyse)
  - Medizin, Medizintechnik
  - Geologie, Ökologie, …
- Ingenieurswissenschaften
  - Automatisierungstechnik,
  - Regelungs- und Steuerungstechnik,
  - Fahrzeugtechnik,
  - Qualitätssicherung (unscharfe Datenanalyse),
  - Unterhaltungselektronik,
  - …

Fuzzy-Mengen (1)

Karl ist 190cm groß

präzises Wissen

Karl ist zwischen 185 und 195cm groß

ist bereits unpräzise, aber die Bereichsgrenzen sind scharf angegeben.

Karl ist groß

vages Wissen (*fuzzy knowledge*),

man kann evtl. einige Werte ausschließen,
wie 150cm, was ist aber mit 1,90m? Möglich …

Fuzzy-Mengen (2)

Gegeben sei eine linguistische Variable X,
dann sei definiert eine Zugehörigkeitsfunktion
(*membership function*) über dem vagen Begriff F μ_F als
($x \in X$): $\mu_F: X \rightarrow [0,1]$ mit

- $\mu_F(x) = 1$ als $x \in F$ „stimmt“
- $\mu_F(x) = 0$ als $x \in F$ „stimmt nicht“
- $0 < \mu_F(x) < 1$ als $x \in F$ „stimmt mit Grad $\mu_F(x)$“

Operation wie bei „scharfen“ Mengen sind möglich, bspw. Durchschnitt, Vereinigung etc.

Fuzzy-Mengen

Beispiel(1)

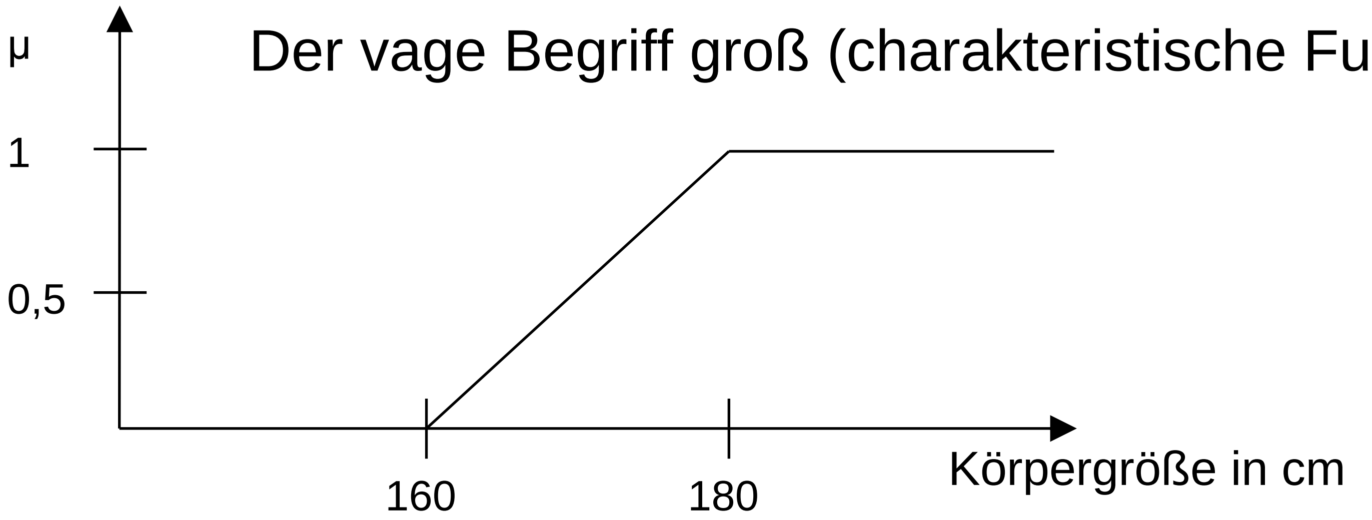
Linguistische Variable: groß

- Festlegung des Wertebereiches für Körpergröße bspw. X= {0,1,…., 280)
- Bestimmen der Zugehörigkeitsfunktion $\mu_{\text{groß}}$
    - $\mu_{\text{groß}}: X \rightarrow [0,1]$

$\mu_{\text{groß}}(x)$ mit $x \in X$ gibt die Zugehörigkeit des Wertes zum Begriff „groß“ an

$\mu_{\text{groß}}(x) = 0$	x gehört definitiv nicht dazu
$\mu_{\text{groß}}(x) = 1$	x gehört definitiv dazu
$0 < \mu_{\text{groß}}(x)<1$ und $\mu_{\text{groß}}(x)=r$	x gehört dazu mit Grad r

Fuzzy-Mengen



- der präzise Begriff B=190cm groß:

$$\mu_B(x) = \begin{cases} 1 \; falls \; x = 190 \\ 0 \; sonst \end{cases}$$

- der unpräzise Begriff B= „zwischen 185 und 195cm groß“

$$\mu_B(x) = \begin{cases} 1 \; falls \; 185 \; \leq x \leq 195 \\ 0 \; sonst \end{cases}$$

150cm gehört nicht zu groß

170cm gehört zum Grad 0.5 dazu

Fuzzy-Mengen

Fuzzy-Mengen werden durch Zugehörigkeitsfunktionen repräsentiert. Die Art der Darstellung ist von der Grundmenge abhängig.

Es gibt 3 verschiedene Darstellungsmöglichkeiten für unscharfe Mengen:

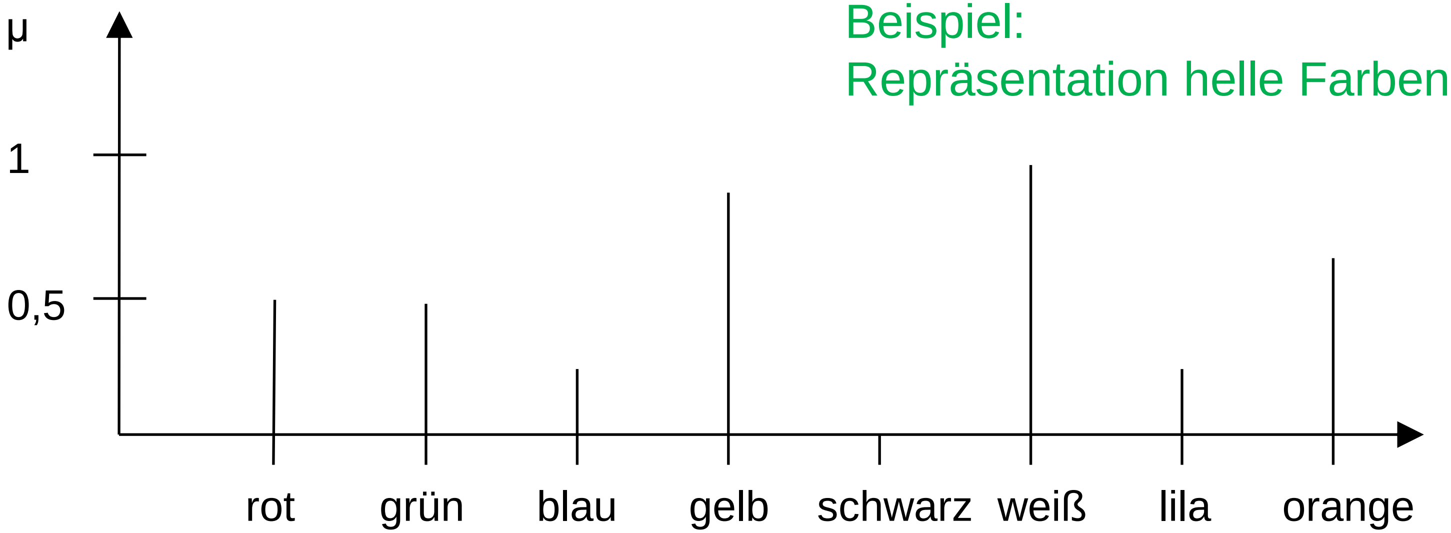
- diskrete Darstellung,
- parametrisierte Darstellung,
- Alphaschnitte.

Fuzzy-Mengen

- diskrete Darstellung:
  der jeweilige x-Wert wird mit dem Zugehörigkeitswert abgelegt. Dazu darf die Grundmenge nur eine endliche Anzahl von Elementen enthalten.

Wenig praktikabel und nur für diskrete Mengen mit wenigen Elementen!

Beispiel:
Repräsentation helle Farben

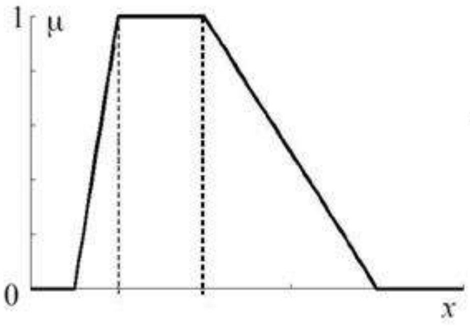


Fuzzy-Mengen

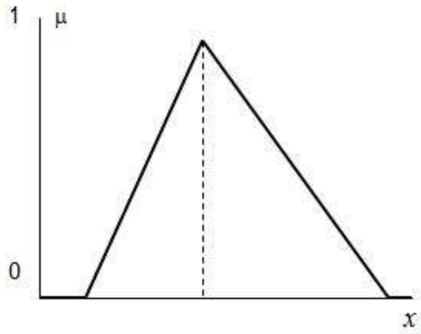
Zugehörigkeitsfunktionen

Häufig auftretende Typen von Zugehörigkeitsfunktionen

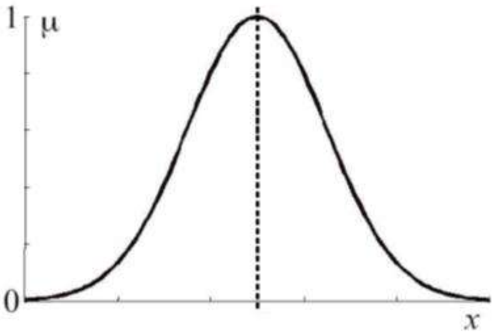
Trapezförmige



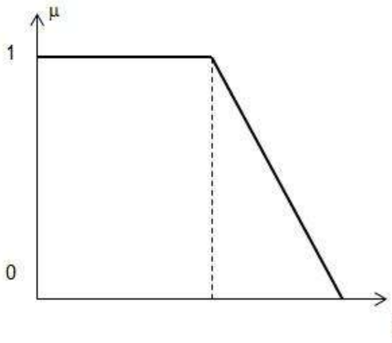
Dreieckförmige



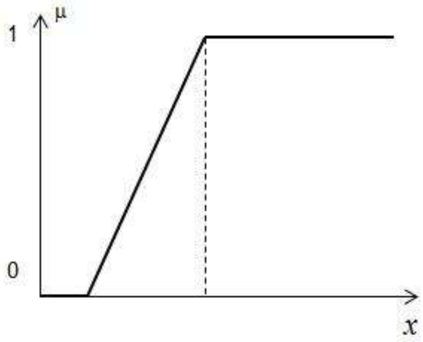
Funktion des Exponentialtyps



monoton fallende Funktion



monoton steigende Funktion



Fuzzy-Mengen parametrisierte Darstellung (1)

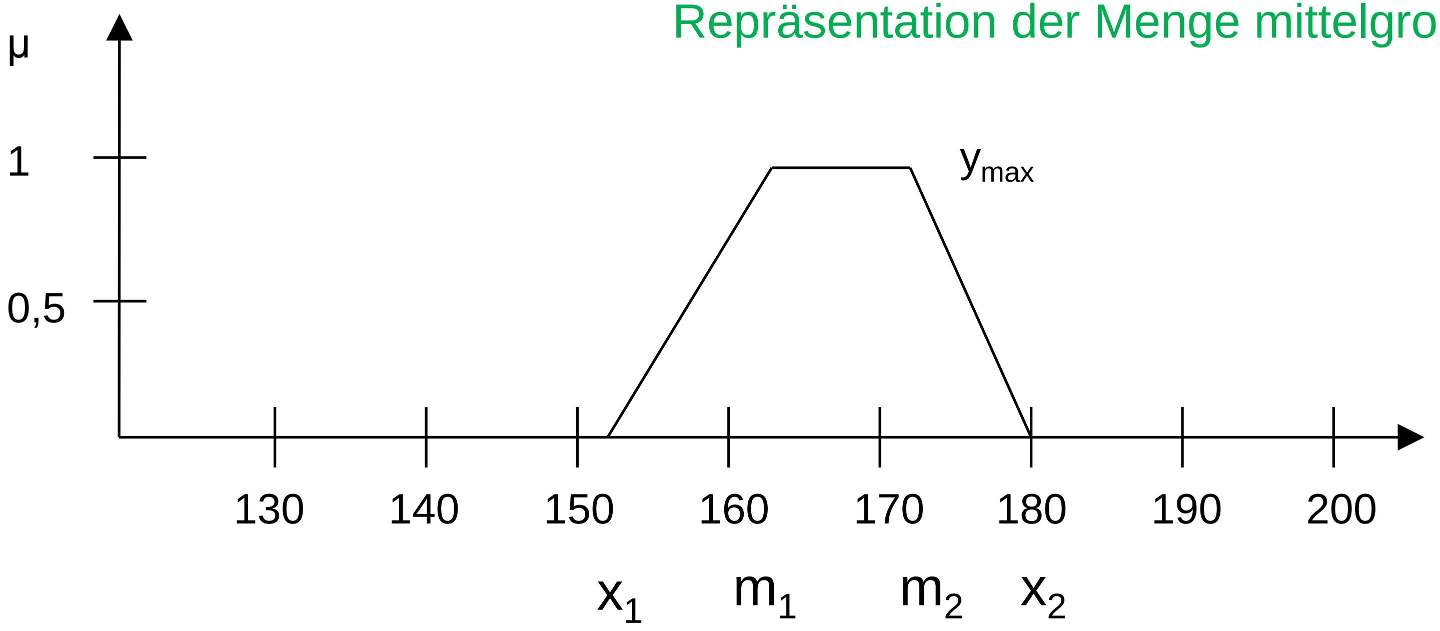
- parametrisierte Darstellung:
  effiziente Darstellung mittels LR-Darstellung
  nützlich, wenn Anzahl der Elemente sehr groß, abzählbar unendlich oder kontinuierlich ist
    - m_1 und m_2 Grenzen des Intervalls mit maximaler Zugehörigkeit
    - L und R als zwei die Gestalt der linken bzw. rechten Flanke der Menge beschreibende Funktionen
    - α und β als die Maße für die Breite der linken bzw. rechten Flanke der Menge
    - mit zusätzlichem Parameter y_{max} maximale Zugehörigkeitswerte <1 möglich

Fuzzy-Mengen

parametrisierte Darstellung (2)

Aus Gründen der Anschaulichkeit hier nicht die absolute Angabe der Flankenbreite in Form von α und β, sondern relative Angabe in Form von Punkten auf der x-Achse ($x_1=m_1-\alpha$ und $x_2=m_2+\beta$)

Beispiel:
Repräsentation der Menge mittelgroßer Männer



Fuzzy-Mengen

parametrisierte Darstellung (3)

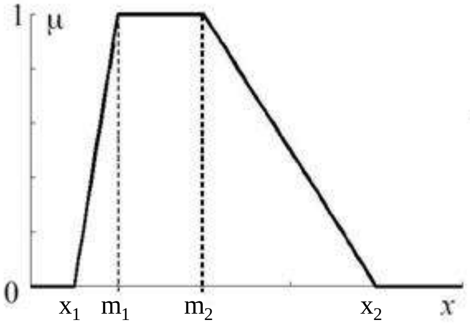
Unscharfe Zahlen bzw. Intervalle sind dann folgendermaßen definiert:

$\tilde{I} := (m_1, m_2, x_1, x_2, y_{max})_{LR}$ mit (für ein Trapez)

$$\mu_{\tilde{I}}(x) := \left\{ \begin{array}{l} \left.\begin{array}{ll} 0 & \text{wenn } x<x_1 \\ \frac{x-x_1}{m_1-x_1} * y_{max} & \text{wenn } x_1 \leq x < m_1 \end{array}\right\} L \\ \begin{array}{ll} y_{max} & \text{wenn } m_1 \leq x \leq m_2 \end{array} \\ \left.\begin{array}{ll} \frac{x_2-x}{x_2-m_2} * y_{max} & \text{wenn } m_2 < x \leq x_2 \\ 0 & \text{wenn } x>x_2 \end{array}\right\} R \end{array} \right.$$

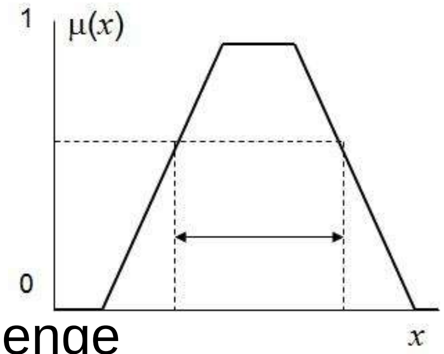
Vorteil:
einfache und effiziente Verarbeitung

Nachteil:
beschränkte Form der darstellbaren Mengen (was sich insbesondere in Näherungsformeln für Multiplikation und Division ausdrückt)

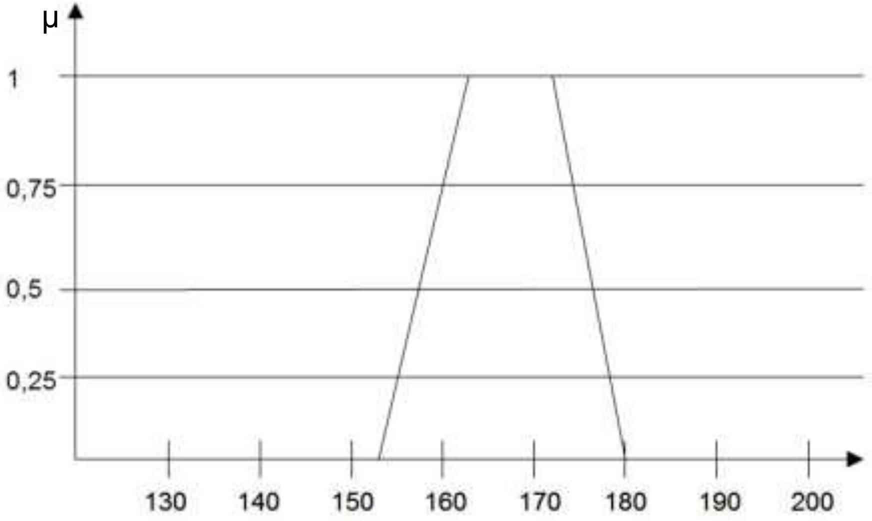


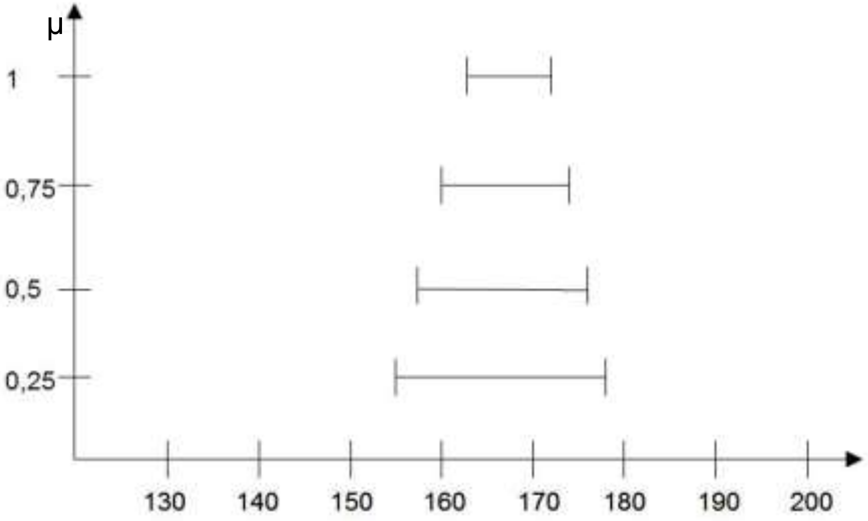
Fuzzy-Mengen

Alphaschnitte



- Darstellung durch Alphaschnitte:
  - horizontale Interpretation parallel zur x-Achse,
  - mehrere Alphaschnitte ein und derselben Menge werden zu einem Mengensystem zusammengefasst,
  - Umso genauer, je mehr Schnitte, dadurch aber auch steigender Verarbeitungsaufwand.

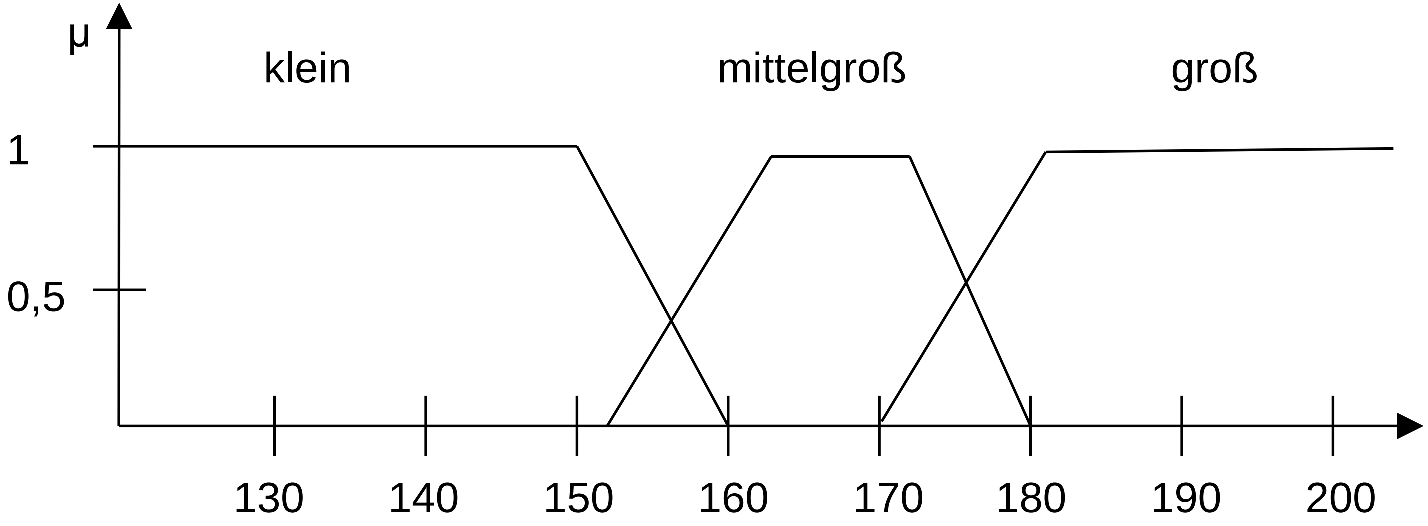




Linguistische Variablen als Fuzzy-Mengen (1)

Beispiel:

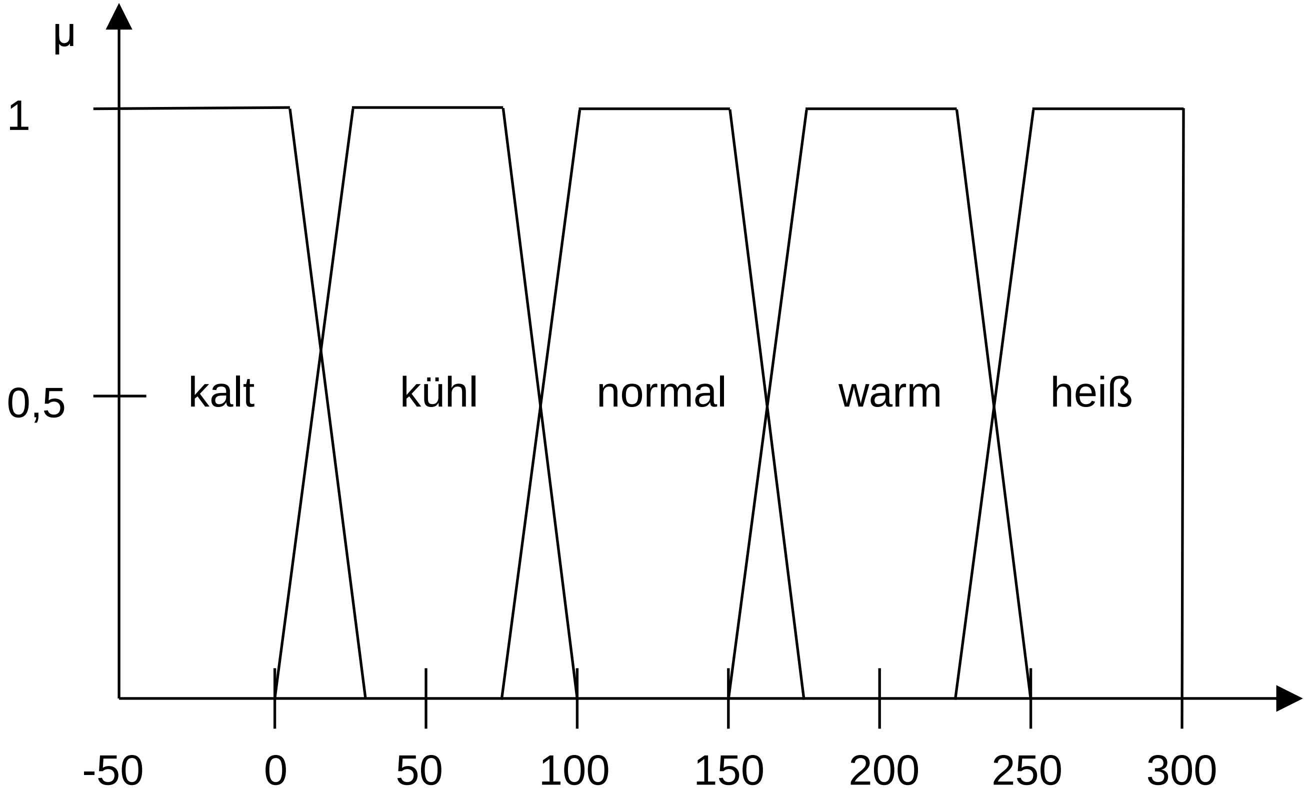
Repräsentation der linguistischen Variable „Körpergröße von Männern“ mit 3 Termen mittels Fuzzy-Mengen



Linguistische Variablen als Fuzzy-Mengen (2)

Beispiel:

Repräsentation der linguistischen Variablen „Lagergehäuse-temperatur von Gleitlagern“ mit 5 Termen mittels Fuzzy-Mengen



Rechnen mit Fuzzy-Mengen

- Ich suche eine Wohnung und die soll möglichst preiswert, entweder stadtnah oder ÖPNV-technisch gut angebunden sein und eine Größe zwischen 90 und 130qm haben.
- Die Lagergehäusetemperatur von Gleitlagern soll normal oder warm sein.

Voraussetzung:

benötigt wird eine **endliche Repräsentation** der Fuzzy-Mengen, z.B.

- Referenzmenge X ist endlich.
- Die Funktion μ ist als analytischer Ausdruck (parameterabhängig) bekannt.

Rechnen mit Fuzzy-Mengen

Teilmenge:

- Seien μ und μ‘ zwei Fuzzy-Mengen von X
- μ heißt **Teilmenge** von μ‘ ($\subset$) **gdw.**
  für alle x $\in$ X gilt: μ(x) $\leq$ μ‘(x).

Rechnen mit Fuzzy-Mengen

Durchschnitt (1)

Realisierung der Verknüpfung von Fuzzy-Mengen:

Durchschnitt:

- Fuzzy-Mengen sind Abbildungen auf [0, 1].
- Durchschnitt von 2 Fuzzy-Mengen reduziert sich auf das elementweise Berechnen einer Funktion: $S : [0, 1] \times [0, 1] \rightarrow [0, 1]$, so dass gilt:
  $(\mu \cap \mu')(x) = S(\mu(x), \mu'(x))$

Welche Eigenschaften muss S erfüllen?

- $S(a, 1) = a$
- Monotonie: $a \leq b \rightarrow S(a, c) \leq S(b, c)$
- Kommutativität: $S(a, b) = S(b, a)$
- Assoziativität: $S(a, S(b, c)) = S(S(a, b), c)$

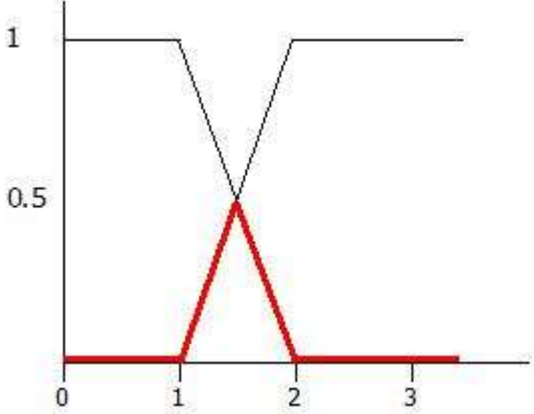
Eine solche Funktion wird auch **t-Norm** genannt.

Rechnen mit Fuzzy-Mengen

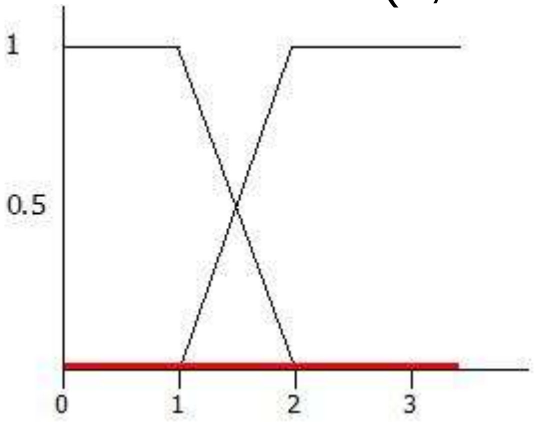
Durchschnitt (2)

Folgende Funktionen erfüllen die Forderungen:

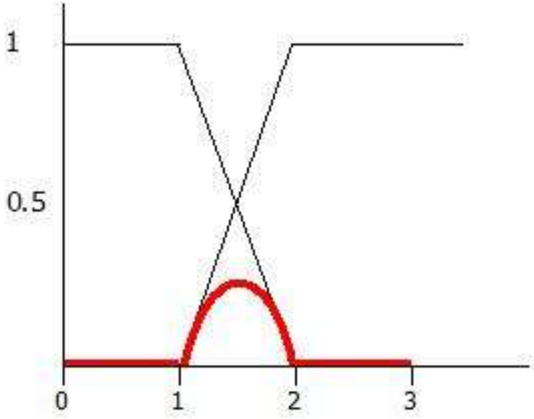
Variante 1: min(a, b)



Variante 2: max(0, a+b-1)



Variante 3: a*b



Rechnen mit Fuzzy-Mengen

Vereinigung (1)

Realisierung der Verknüpfung von Fuzzy-Mengen:

Vereinigung:

- Fuzzy-Mengen sind Abbildungen auf [0, 1].
- Vereinigung von 2 Fuzzy-Mengen reduziert sich auf das elementweise Berechnen einer Funktion: $V : [0, 1] \times [0, 1] \rightarrow [0, 1]$, so dass gilt:
  $(\mu \cup \mu')(x) = V(\mu(x), \mu'(x))$

Welche Eigenschaften muss V erfüllen?

- $V(a, 0) = a$
- Monotonie: $a \leq b \rightarrow V(a, c) \leq V(b, c)$
- Kommutativität
- Assoziativität

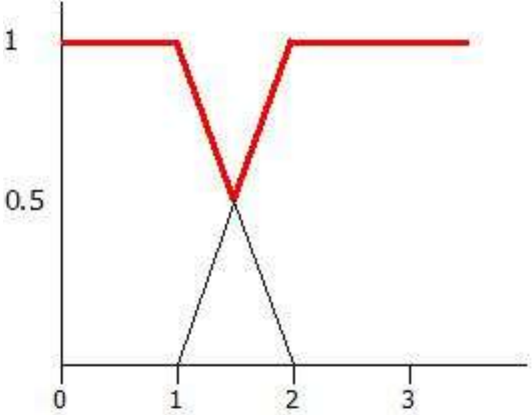
Eine solche Funktion wird auch **t-CoNorm** genannt.

Rechnen mit Fuzzy-Mengen

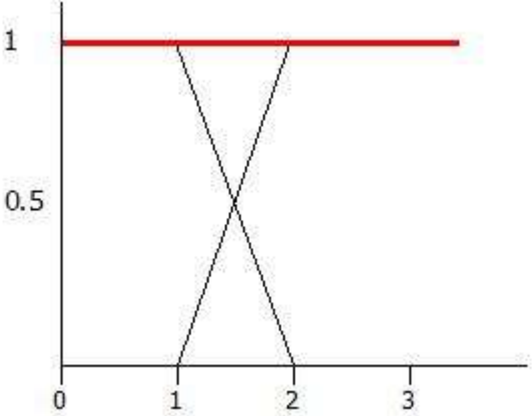
Vereinigung (2)

Folgende Funktionen erfüllen die Forderungen:

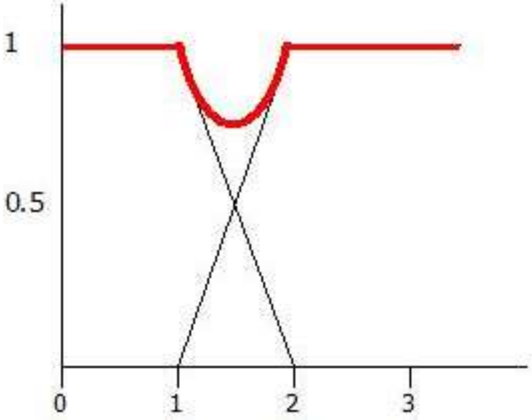
Variante 1: max(a, b)



Variante 2: min(1, a+b)



Variante 3: a+b-a*b



Rechnen mit Fuzzy-Mengen

Disjunktion, Konjunktion

- Es gibt noch weitere Varianten für Konjunktion und Disjunktion.
- Welche der Varianten (1, 2 oder 3) benutzt wird, ist Ihnen überlassen.
- Die Varianten liefern i.a. unterschiedliche Resultate, so dass es nützlich ist, eine Entscheidungsstrategie zu entwickeln.
  Z.Bsp. könnte man mit allen 3 Varianten rechnen und nimmt dann das Angebot, welches am häufigsten gewinnt.
- Achtung: die gewählten Funktionen für Durchschnitt und Vereinigung müssen zueinander passen, um in der klassischen Logik geltende Aussagen auch hier zu erfüllen.
  Wählt man
  - für den Durchschnitt das Minimum,
  - für die Vereinigung das Maximum,

  } Variante 1

  so gilt auch in der Fuzzy-Logik die Aussage: $A \vee (A \wedge B) = A$
  Denn es gilt:
  $\mu_{A \vee (A \wedge B)}(x) = \max(\mu_A(x), \min(\mu_A(x), \mu_B(x))) = \mu_A(x)$

Rechnen mit Fuzzy-Mengen

Negation, Implikation

Komplementbildung:

- $\mu_{Zc}(x) = 1- \mu_Z(x)$

Implikation:

Die Varianten der Definition der Implikation haben natürlich etwas mit den 3 Varianten für $\wedge$ und $\vee$ zu tun.

Die Implikation kann man z.Bsp. folgendermaßen definieren:

Name	$\mu_{A\Rightarrow B}(x)$
Łukasiewicz	$\min(1, 1 - \mu_A(x) + \mu_B(x))$
Kleene-Dienes	$\max(1 - \mu_A(x), \mu_B(x))$
Zadeh	$\max(1 - \mu_A(x), \min(\mu_A(x), \mu_B(x)))$
Reichenbach	$1 - \mu_A(x) + \mu_A(x) * \mu_B(x)$

Rechnen mit Fuzzy-Mengen

Beispiel (1)

Alice sucht eine Wohnung und die soll möglichst preiswert, entweder stadtnah oder ÖPNV-technisch gut angebunden sein und eine Größe zwischen 90 und 130qm haben.

- 3 Angebote:

	Fläche	Stadtnähe (km)	Preis (€)	Bus/h
Wohnung 1	120	5	850	3
Wohnung 2	154	7	900	2
Wohnung 3	85	4	700	4

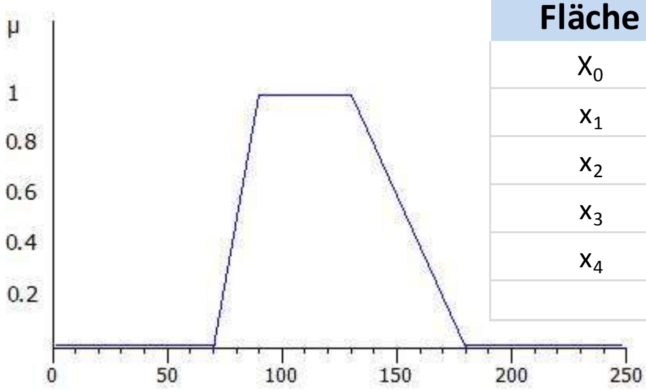
- Wir haben 4 linguistische Variablen.
- Diese müssen mit
  - oder:
    stadtnah **oder** gute ÖPNV-Anbindung bzw.
  - und:
    preiswert **und** 90-130qm **und**
    (stadtnah oder gute ÖPNV-Anbindung)

verbunden werden.

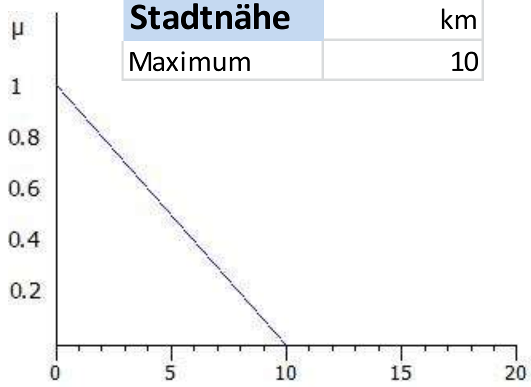
Rechnen mit Fuzzy-Mengen Beispiel (2)

1. Fuzzymengen für die unscharfen Begriffe bilden. Beispielsweise:

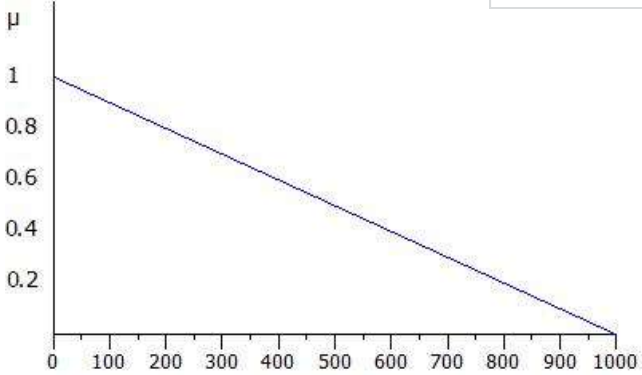
Fläche	m^2		Anstieg
X_0	0	0	0,0000
x_1	70	0	0,05
x_2	90	1	0,0000
x_3	130	1	-0,02
x_4	180	0	0,0000
	200	0	



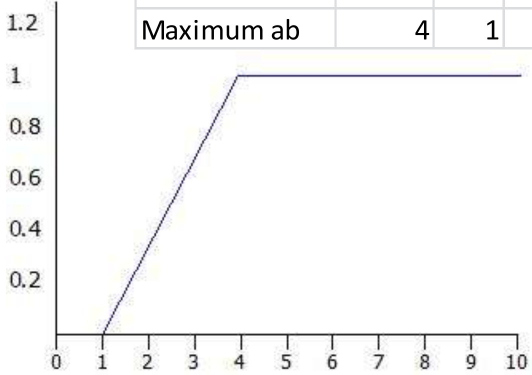
Stadtnähe	km
Maximum	10



Preis	€
Maximum	1000



ÖPNV	Bus/h		Anstieg
Minimum	1	0	0,33
Maximum ab	4	1	



Rechnen mit Fuzzy-Mengen

Beispiel(3)

2. Und nun die günstigste Wohnung aussuchen.

	Real				Fuzzy			
	Fläche	Preis	Stadtnähe	ÖPNV	Fläche	Preis	Stadtnähe	ÖPNV
Wohnung1	120	850	5	3	1	0,15	0,5	0,67
Wohnung2	154	900	7	2	0,52	0,1	0,3	0,33
Wohnung3	85	700	4	4	0,75	0,3	0,6	1

Fläche ∧ Preis ∧ (Stadtnähe ∨ ÖPNV)

- Variante 1:
  und: min(a,b)
  oder: max(a,b)
- Variante 2:
  und: max(0,a+b-1)
  oder: min(1,a+b)
- Variante 3:
  und: a*b
  oder: a+b-a*b

	Stadtnähe ∨ ÖPNV			Fläche ∧ Preis ∧ (Stadtnähe ∨ ÖPNV)			Rang		
	Variante 1	Variante 2	Variante 3	Variante 1	Variante 2	Variante 3	Variante 1	Variante 2	Variante 3
Wohnung1	0,67	1	0,83	0,15	0,15	0,125	2	1	2
Wohnung2	0,33	0,63	0,53	0,1	0	0,028	3	3	3
Wohnung3	1	1	1	0,3	0,05	0,225	1	2	1

Fuzzy-Control

Fuzzy: Anwendung für Regelungsvorgänge

Wenn die Unwucht groß ist und die Drehzahl der Trommel hoch ist, dann Geschwindigkeit stark drosseln.

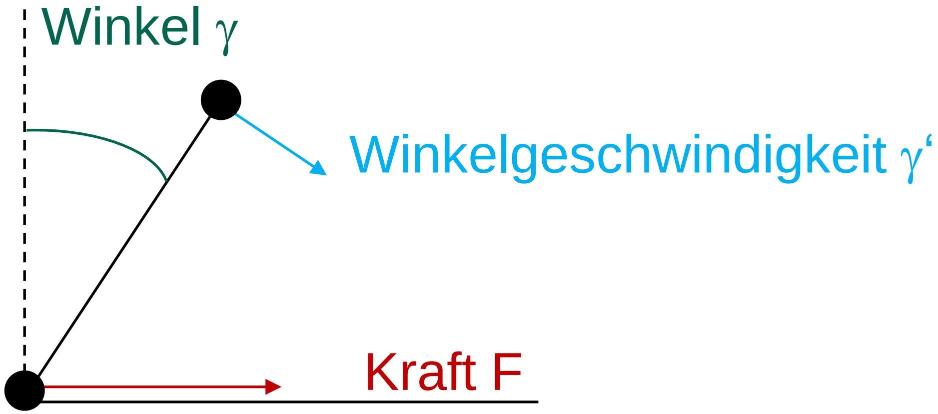
groß, *hoch* und *stark* sind unscharfe Begriffe und können bspw. als Fuzzy-Mengen modelliert werden.

Regelaufbau:

- linke Seite:
  Messgrößen $\varphi_1 \in X_1, \ldots, \varphi_n \in X_n$
- rechte Seite:
  Einstellgrößen,
  nachfolgend gehen wir davon aus, dass rechts nur **EINE** Einstellgröße steht $\psi \in Y$
- Regeln der Form:
  R_j: WENN φ_1 IST $a_{1,j} \wedge \ldots \wedge \varphi_n$ IST $a_{n,j}$ DANN ψ IST b_j

Inverses Pendel – Stabbalance Problem

vertikales Balancieren eines Stabes, nur das untere Stabende darf bewegt werden



Messgrößen:

- X_1 … Winkel γ: [-90, +90] Grad
- X_2 … Winkelgeschwindigkeit γ': [-45, +45] Grad/Sekunde

Einstellgröße:

- Y … (horizontale) Kraft F: [-10, +10] Newton

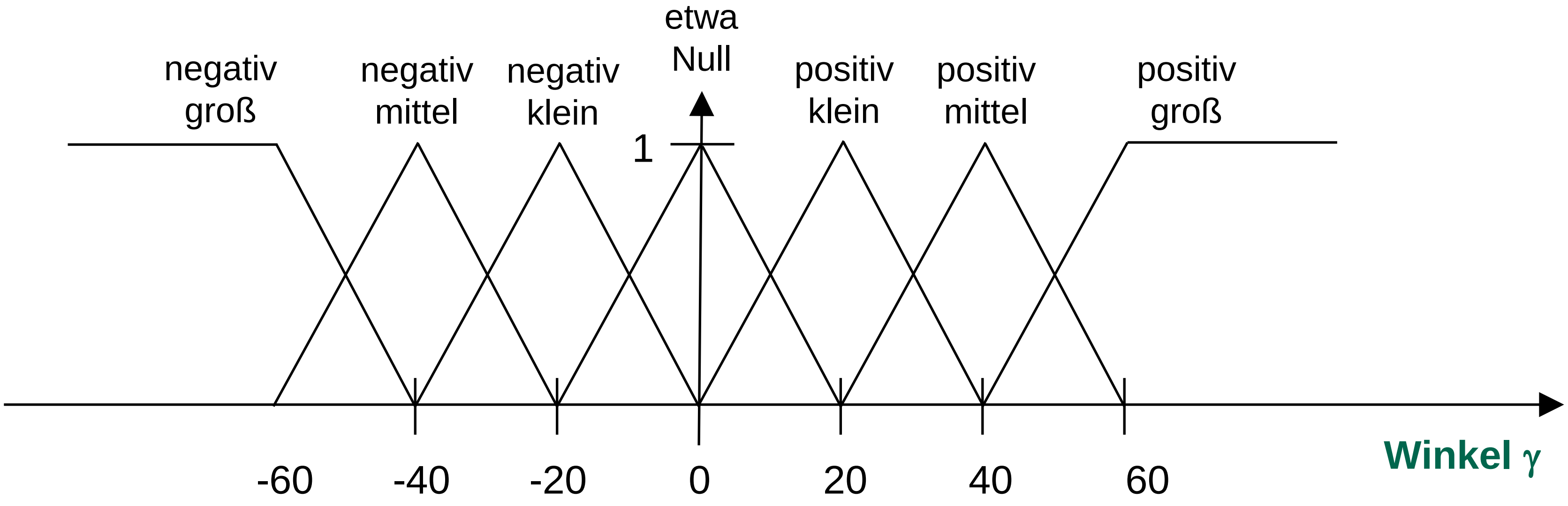
Fuzzy-Control

Inverses Pendel (2)

WENN γ IST positiv klein $\wedge$ γ' ist etwa Null

DANN F IST positiv klein

Die Werte *positiv klein*, *etwa Null* etc. kann man mittels Fuzzy-Dreiecksfunktionen modellieren, nachfolgend Beispielmodellierung für den **Winkel**:



Die Zuordnung eines Messwertes zu den Fuzzy-Mengen bezeichnet man als Fuzzyfizierung.

Inverses Pendel (3)

Regelwerk:

γ

		ng	nm	nk	en	pk	pm	pg
	ng				ng			
	nm				nm			
	nk			nm	nk	en		
γ‘	en	ng	**nm**	nk	en	pk	pm	pg
	pk			en	pk	pm		
	pm				pm			
	pg				pg			

WENN γ IST negativ mittel und γ‘ IST etwa Null
DANN F IST negativ mittel.

Inverses Pendel (4)

Verarbeitung:

Welche Regel ist anzuwenden? Meistens kommen mehrere Regeln in Betracht.

Für die konkreten Messwerte

- werden für jede Bedingung jeder Regel der Zugehörigkeitswert bestimmt und
- die Werte für die Bedingungen werden je Regel mit *min* verknüpft.

Regeln mit einem Wert 0 treffen nicht zu,

Regeln mit einem Wert > 0 treffen mehr oder minder zu.

Für die Auswahl gibt es nun 2 Möglichkeiten:

- Regel mit dem maximalen Wert für die Bedingung (*max*)
- alle Regeln werden berücksichtigt, deren Wert > 0 ist

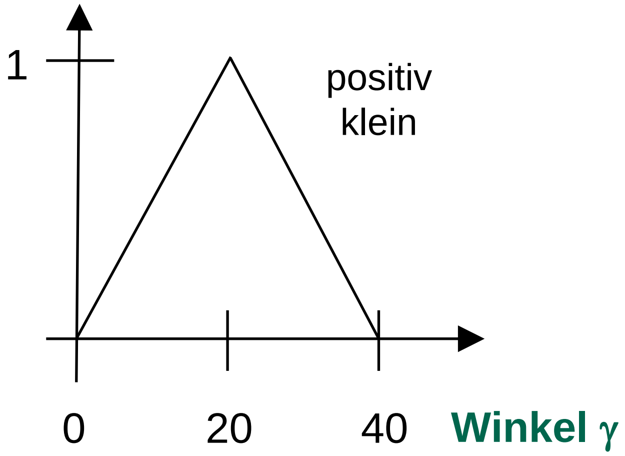
Fuzzy-Control

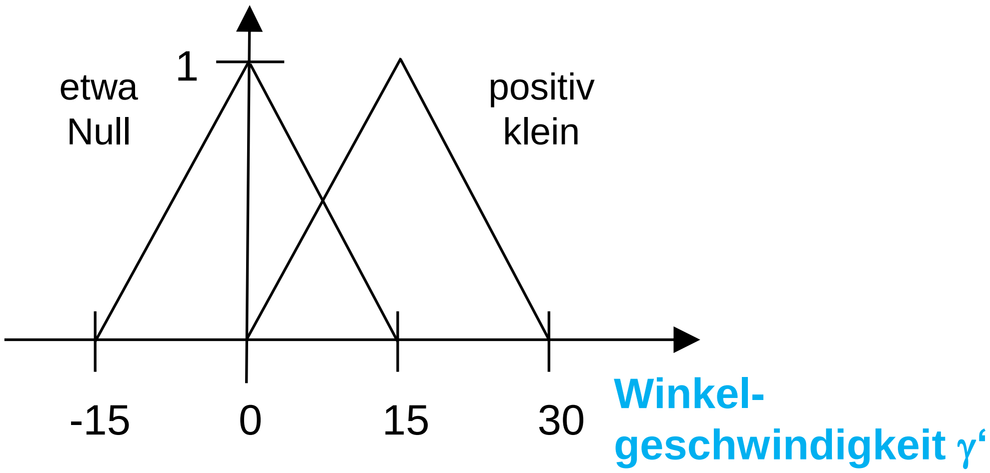
Inverses Pendel Beispiel (1)

Fallbeispiel:

γ = 20 Grad

γ‘ = 10 Grad/Sekunde





Alle Regeln anwendbar mit γ ist *positiv klein* und γ‘ ist *etwa Null* bzw. *positiv klein*.

Es kommen 2 Regeln in Frage:

① Wenn γ ist *positiv klein* und γ‘ ist *etwa Null* dann F ist positiv klein.

② Wenn γ ist *positiv klein* und γ‘ ist *positiv klein* dann F ist positiv mittel.

Inverses Pendel Beispiel (2)

Zugehörigkeitswerte der Regeln ergibt sich aus dem Minimum:

① Wenn γ ist *positiv klein* und γ‘ ist *etwa Null*
dann F ist positiv klein.
$\min(1,{}^1/_3) = {}^1/_3$

② Wenn γ ist *positiv klein* und γ‘ ist *positiv klein*
dann F ist positiv mittel.
$\min(1,{}^2/_3) = {}^2/_3$

- Für alle anderen Regeln: Wert=0

Variante Maximum:

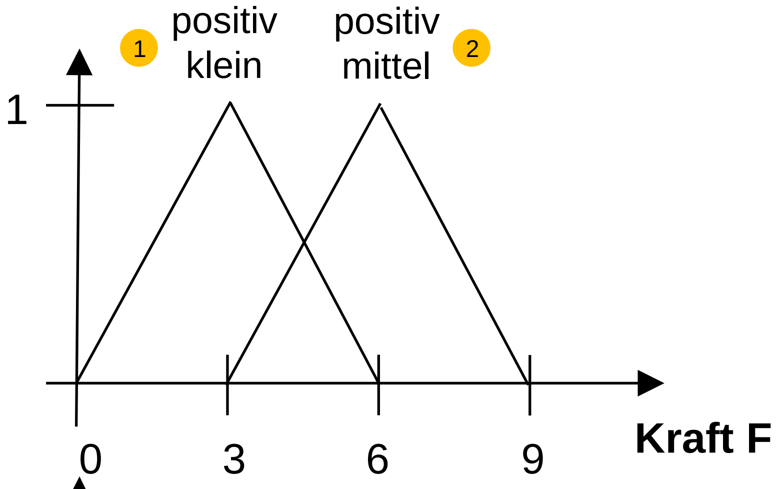
$\max({}^1/_3, {}^2/_3) = {}^2/_3$
Regel „Wenn γ ist *positiv klein* und γ‘ ist *positiv klein*
dann F ist positiv mittel“ wird ausgewählt.

Variante beide Regeln werden berücksichtigt. ☞

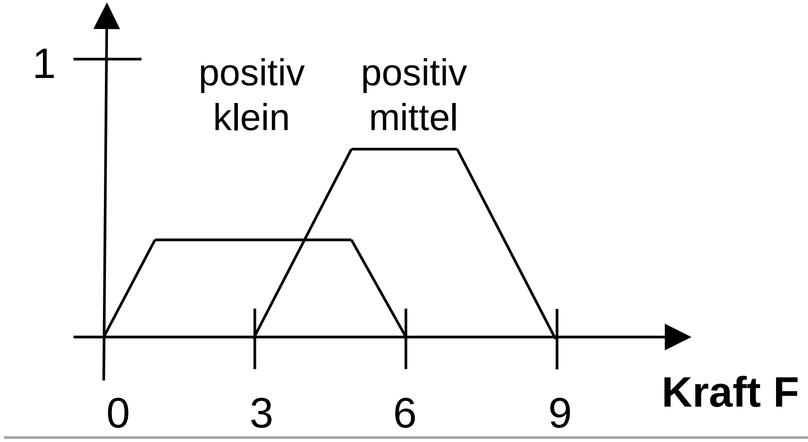
Inverses Pendel Beispiel (3)

Wie wird die anzulegende Kraft F berechnet?

Jede ausgewählte Regel bestimmt eine Fuzzy-Menge für F.



Jede Regel schneidet die von ihr gelieferte Fuzzy-Menge für F ab.



Bei Auswahl der Regel durch das Maximum (Variante 1) ist die resultierende Fuzzy-Menge das abgeschnittene positiv mittel, bei Variante 2 ist die resultierende Fuzzy-Menge die Vereinigung beider Fuzzy-Mengen.

Defuzzyfizierung

Wie ermittelt man nun aus der Fuzzy-Menge für F den konkreten Wert, den man einstellen kann?

Dieses Verfahren nennt man Defuzzyfizierung.

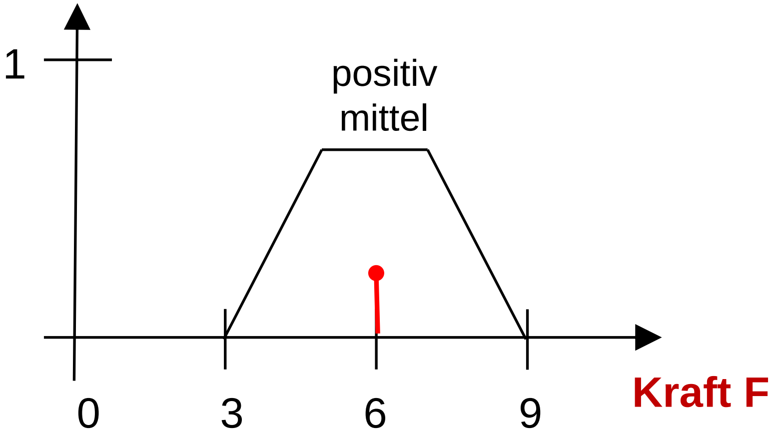
Es existieren verschiedene Verfahren für Defuzzifizierung, wobei 2 unterschiedliche miteinander konkurrierende Ziele verfolgt werden:

- Ermittlung des besten Kompromisses,
  d.h. des x-Wertes, der die „Mitte“ der unscharfen Menge am besten beschreibt,
  z. Bsp. die Schwerpunktmethode (center of area)
- Ermittlung des plausibelsten Wertes,
  d.h. eines x-Wertes mit maximaler Zugehörigkeit μ.

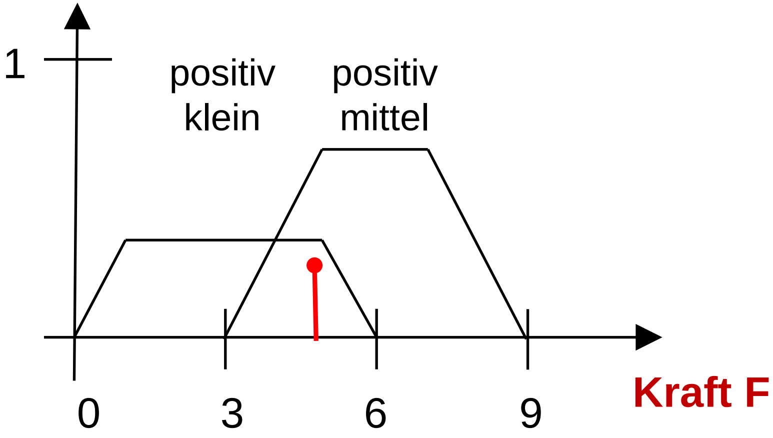
Inverses Pendel Beispiel (4)

Für unser Beispiel wählen wir die Schwerpunktmethode, da sie ein glattes Regelverhalten zur Folge hat.

Der x-Wert unter dem Schwerpunkt wird als Einstellgröße angenommen.

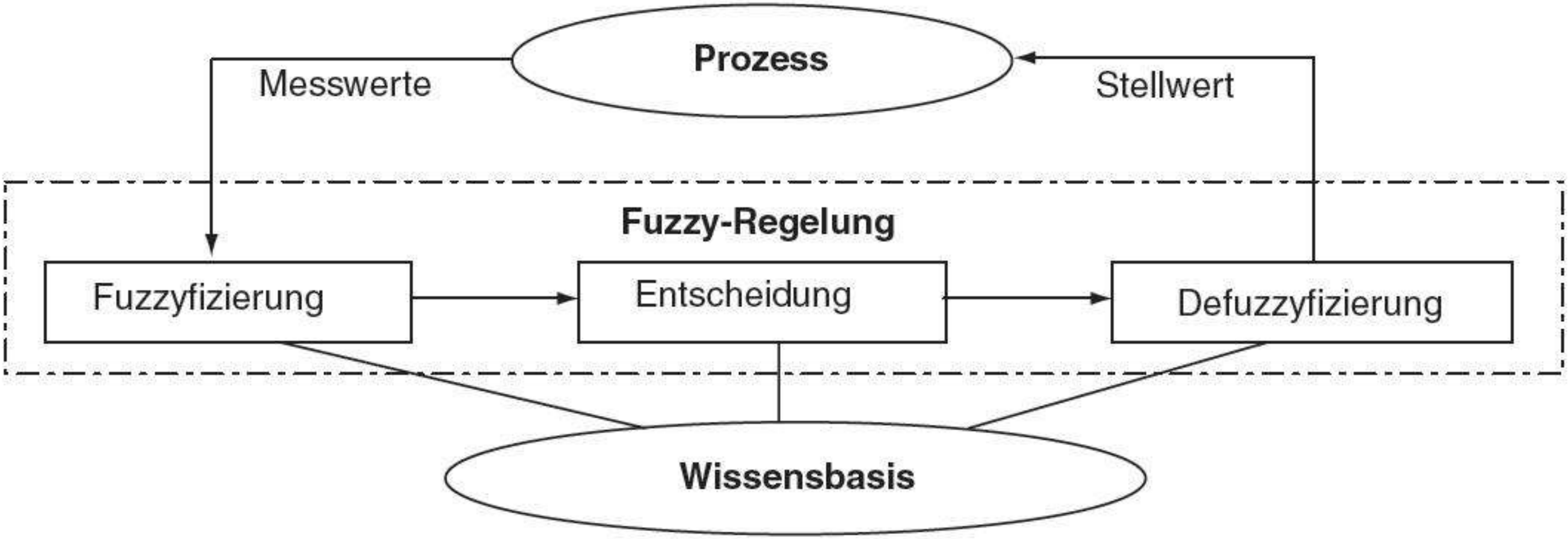


F = 6N



F ≈ 5N

Architektur



Architektur eines Fuzzy-Reglers aus
Uwe Lämmel, Jürgen Cleve: *Lehr- und Übungsbuch Künstliche Intelligenz*. 4. Aufl., Hanser, München, 2012

Lernziele

… Begriffe: Wahrscheinlichkeit, bedingte Wahrscheinlichkeit, abhängige und unabhängige Ereignisse, Induktion und Deduktion, Bayessches Netz, Fuzzyfizierung, Defuzzyfizierung verstanden haben,

… Bayessche Formel anwenden können.

… Fuzzy-Mengen, linguistische Variablen für Probleme erstellen und darstellen können,

… Zugehörigkeitswerte aus Fuzzy-Mengen ermitteln / berechnen können,

… Rechnen mit Fuzzy-Mengen beherrschen,

… Fuzzy-Control verstanden haben.