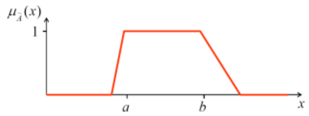
COMPUTATIONAL INTELLIGENCE

11. FUZZY-ARITHMETIK, GRAFISCHE MODELLE

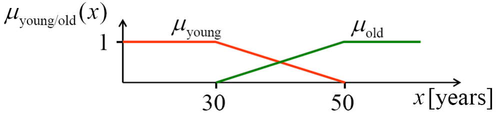
Prof. Dr. Sven Behnke

LETZTE VORLESUNG

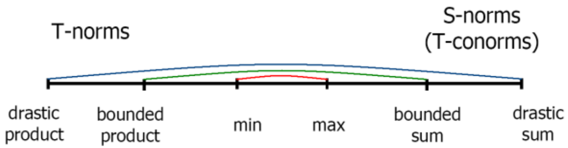
- Fuzzy-Mengen beschrieben durch Mitgliedschaftsfunktion



- Linguistische Variablen und Werte



- Operatoren auf Fuzzy-Mengen
  - Konjunktion (z.B. min)
  - Disjunktion (z.B. max)
  - Negation ($1-\mu_A(x)$)

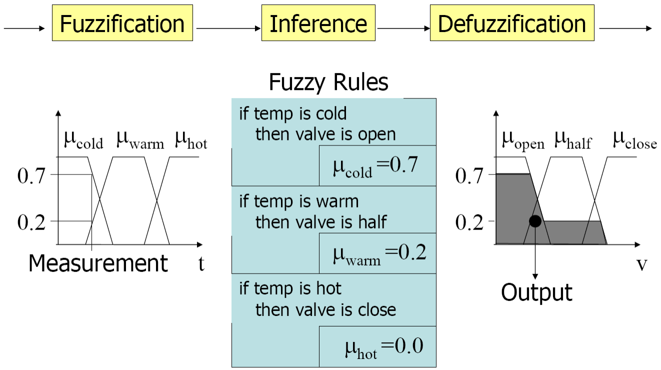


- Fuzzy-Implikation:

$$\mu_{A\to B}(x) := \underbrace{\max}_{*\vee*}\{\underbrace{1-\mu_A(x)}_{\neg A}, \underbrace{\min\{\mu_A(x), \mu_B(x)\}}_{A\wedge B}\}$$

- Fuzzy-Regeln:
  - IF <Prämisse> THEN <Konklusion>
    - Version 1 (Mamdani): Prämisse Konjunktion,
      Konklusion Fuzzy-Menge
    - Version 2 (Takagi-Sugeno): Prämisse Konjunktion,
      Konklusion einfache reelle Funktion

- Fuzzy-Regelsystem:
  - Fuzzyfizierung der Eingaben
  - Aktivierungen der Prämissen
  - Kombination der Konklusionen
  - Defuzzyfizierung

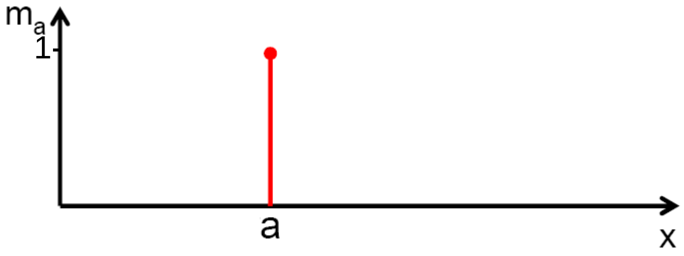


FUZZY-ZAHLEN: MOTIVATION

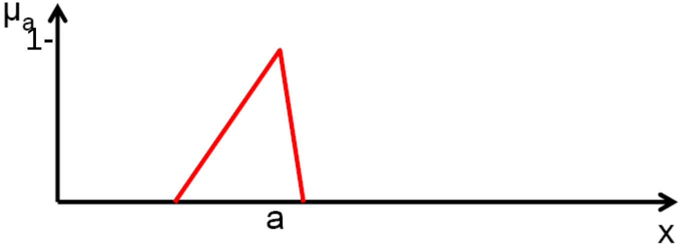
- Häufig ist es erwünscht, ungenaue Eingaben vorzuverarbeiten
  - Addition und Multiplikation von Fuzzy-Zahlen
  - Anwendung von Funktionen (Normalisierung, …) auf Fuzzy-Zahlen
- Fuzzy-Regler können erweitert werden, ungenaue Eingaben in Form von Fuzzy-Zahlen zu akzeptieren
- Fuzzy-Zahlen: Modelliere z.B. Mess-Ungenauigkeit

FUZZY-ZAHLEN

- Klassische “scharfe” Zahlen: $a \in R$
  - Charakteristische Funktion $m_a(x) \in \{0,1\}$



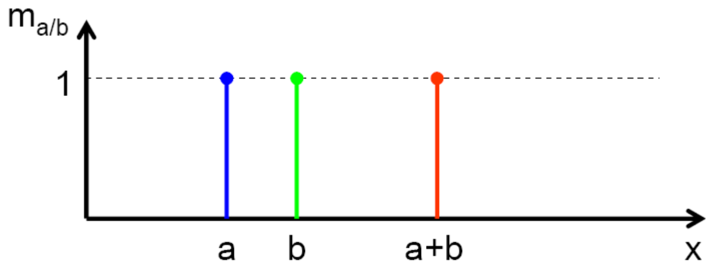
- Fuzzy-Zahlen (“ungefähr a”):
  - Mitgliedschafts-Funktion $\mu_a(x) \in [0,1]$



Mitgliedschafts-Funktionen von Fuzzy-Zahlen sind (üblicherweise):
- **Normalisiert**
- **Monoton** (links/rechts)
- **stückweise ableitbar**
- **genau an einer Stelle 1**

OPERATIONEN AUF FUZZY-ZAHLEN

- Addition zweier scharfer Zahlen:



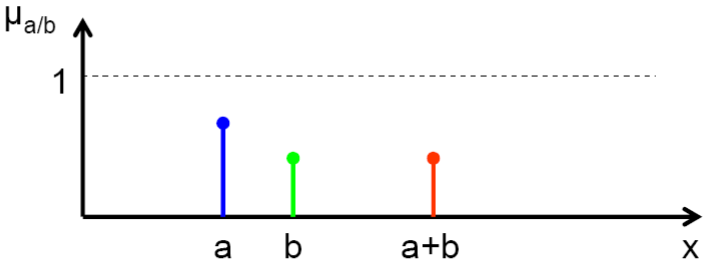
$$m_{a+b}(z) = \begin{cases} 1 & z = a + b \\ 0 & \text{else} \end{cases}$$

$$m_{a+b}(z) = \max_{x,y \in \mathrm{R}} \{ m_a(x) \cdot m_b(y) \mid x + y = z \}$$

= “and”?...

OPERATIONEN AUF FUZZY-ZAHLEN

- Addition zweier Singletons:

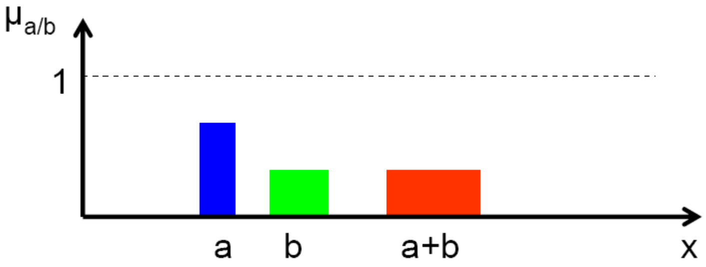


$$\mu_{a+b}(z) = \begin{cases} \min\{\mu_a(a), \mu_b(b)\} & z = a+b \\ 0 & \text{else} \end{cases}$$

$$\mu_{a+b}(z) = \max_{x,y\in R}\{\underbrace{\min\{\mu_a(x), \mu_b(y)\}}_{=\text{ and}} \mid x+y=z\}$$

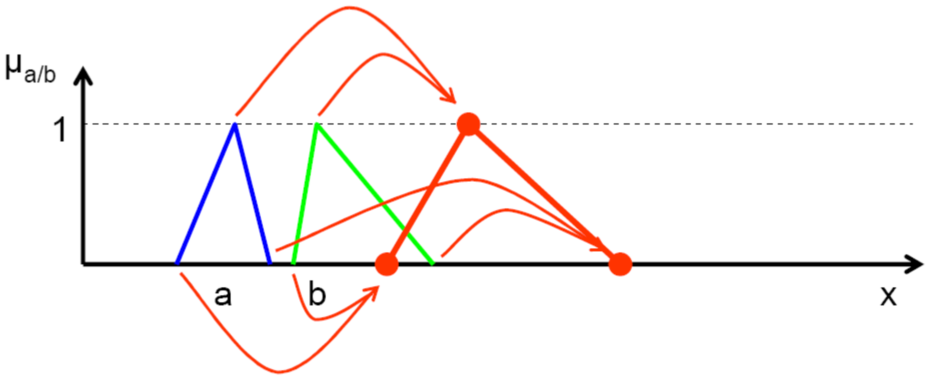
ADDITION ZWEIER FUZZY-INTERVALLE

- Intervall-Arithmetik



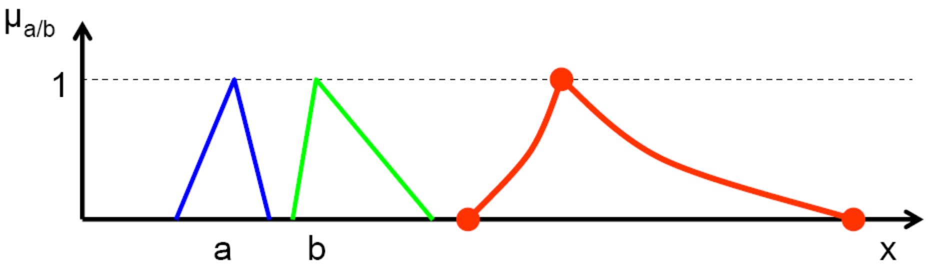
$$\mu_{a+b}(z) = \max_{x,y \in \mathrm{R}} \{ \underbrace{\min\{\mu_a(x), \mu_b(y)\}}_{=\text{ and}} \mid x + y = z \}$$

ADDITION ZWEIER FUZZY-ZAHLEN



$$\mu_{a+b}(z) = \max_{x,y \in \mathrm{R}} \{\min\{\mu_a(x), \mu_b(y)\} \mid x + y = z\}$$

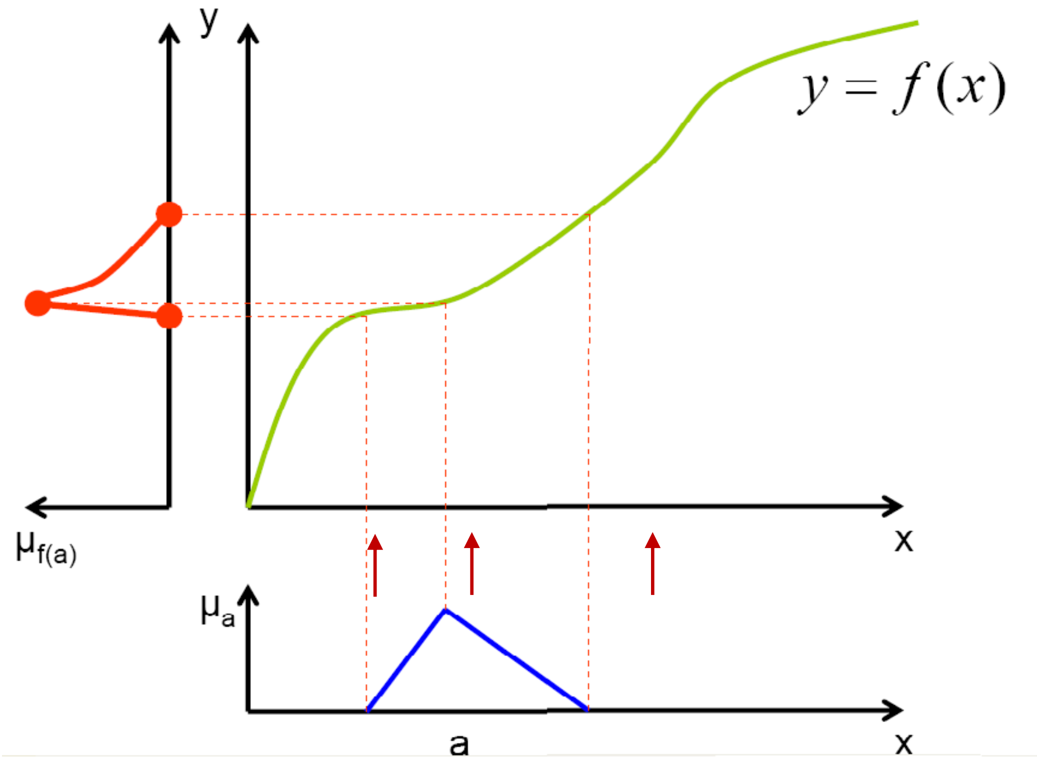
MULTIPLIKATION ZWEIER FUZZY-ZAHLEN



$$\mu_{a*b}(z) = \max_{x,y \in \mathrm{R}} \{\min\{\mu_a(x), \mu_b(y)\} \mid x * y = z\}$$

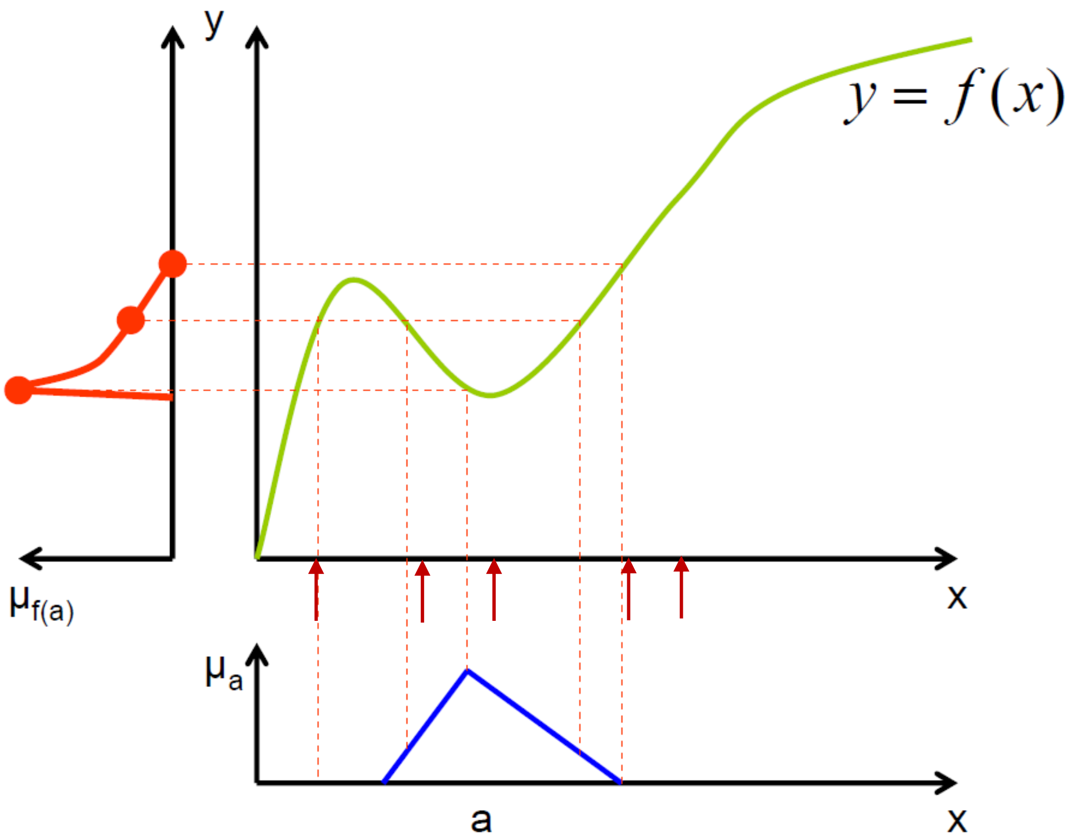
ANWENDUNG SCHARFER FUNKTIONEN AUF FUZZY-ZAHLEN

$$\mu_{f(a)}(y) = \max_{x \in R} \{\mu_a(x) \mid y = f(x)\}$$



ANWENDUNG NICHTMONOTONER FUNKTIONEN AUF FUZZY-ZAHLEN

$$\mu_{f(a)}(y) = \max_{x \in R}\{\mu_a(x) \mid y = f(x)\}$$



ANWENDUNG VON FUNKTIONEN MIT MEHREREN EINGABEN

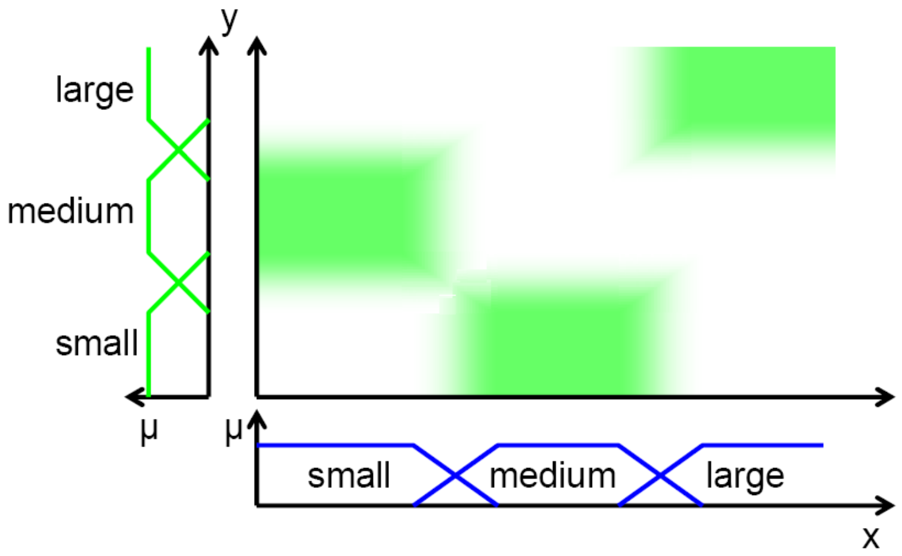
- Nach Erweiterungsprinzip:

$$\mu_{f(a)}(y) = \max_{x_1,\dots,x_n \in R} \{\min\{\mu_{a_1}(x_1),\dots,\mu_{a_n}(x_n)\} \mid y = f(x_1,\dots,x_n)\}$$

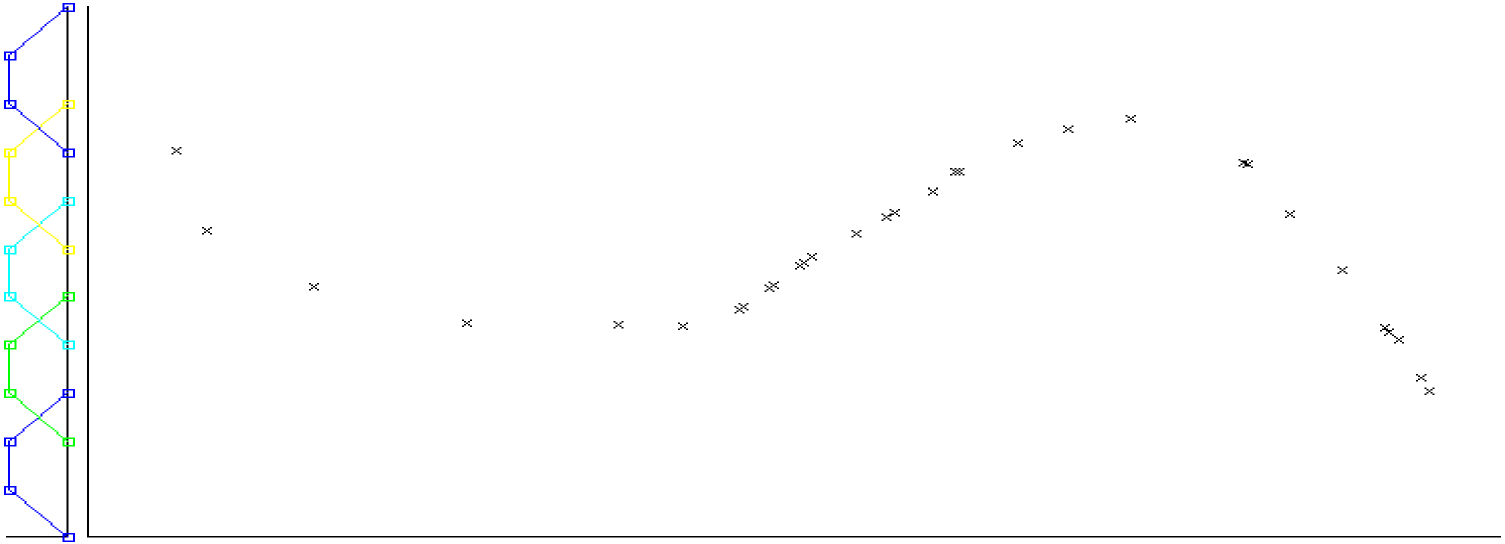
- Berechnung des Maximums unpraktisch
- Approximative Berechnung:
  - Folge von α-Cuts ➔ Rückführung auf Intervall-Arithmetik
  - Polynomielle Repräsentation der linken und der rechten Seite von Fuzzy-Zahlen (abgeschlossen unter Addition und Multiplikation)
  - Diskretisierung der Variablen (Grid-basiert)
  - …

FUZZY-FUNKTIONEN (FUZZY-GRAPHEN)

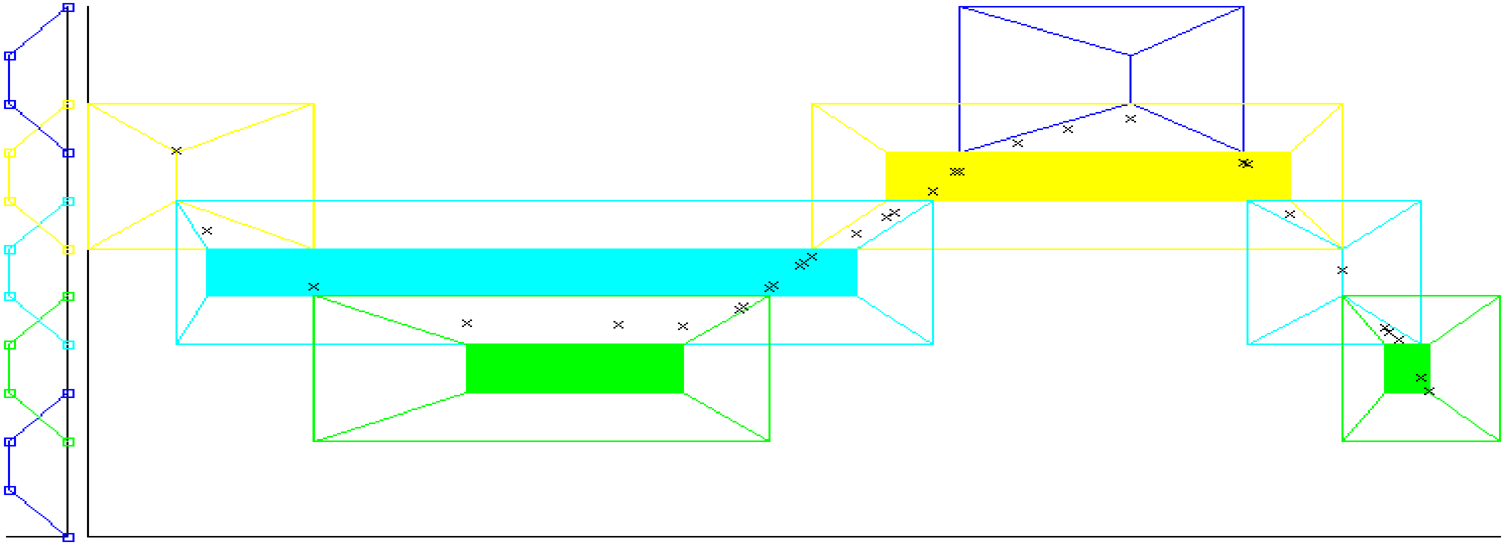
- Reellwertige Funktionen: Unendliche Menge scharfer Punkte
- Fuzzy-Graph: Endliche Menge ungenauer Punkte
  - IF x IS small THEN y IS medium
  - IF x IS medium THEN y IS small
  - IF x IS large THEN y IS large



FUZZY-GRAPH: BEISPIEL



FUZZY-GRAPH: BEISPIEL



FUZZY-FUNKTIONEN (FUZZY-GRAPHEN)

- Jeder Fuzzy-Punkt beschreibt eine ungenaue Relation:

$$(\vec{x}, y) \text{ IS } \mathbf{A} \times B$$

- Ein Fuzzy-Graph ist eine Menge von Fuzzy-Punkten:

$$(\vec{x}, y) \text{ IS } \mathbf{A}_1 \times B_1 \text{ OR} \ldots \text{OR } \mathbf{A}_r \times B_r$$

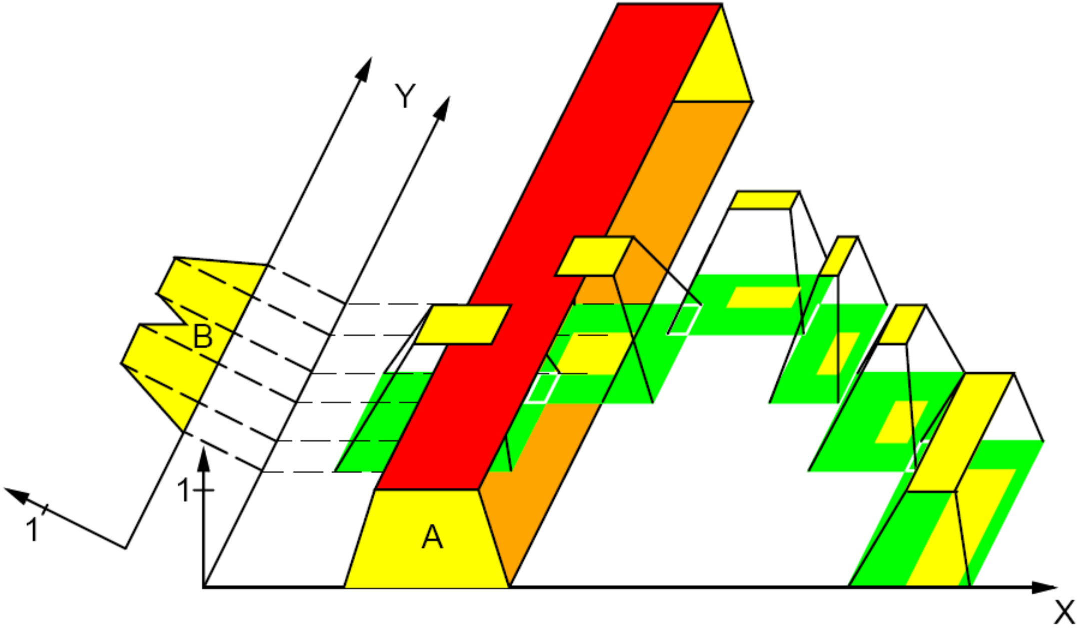
- Inferenz mit Fuzzy-Graphen:

$$\begin{array}{l} \vec{x} \text{ IS } \mathbf{A} \\ \tilde{f} \text{ IS } \bigcup_{\mathrm{j}=1}^{\mathrm{r}} \mathbf{A}_j \times B_j \\ \hline y \text{ IS } B \end{array}$$

BERECHNUNG VON FUZZY-GRAPHEN

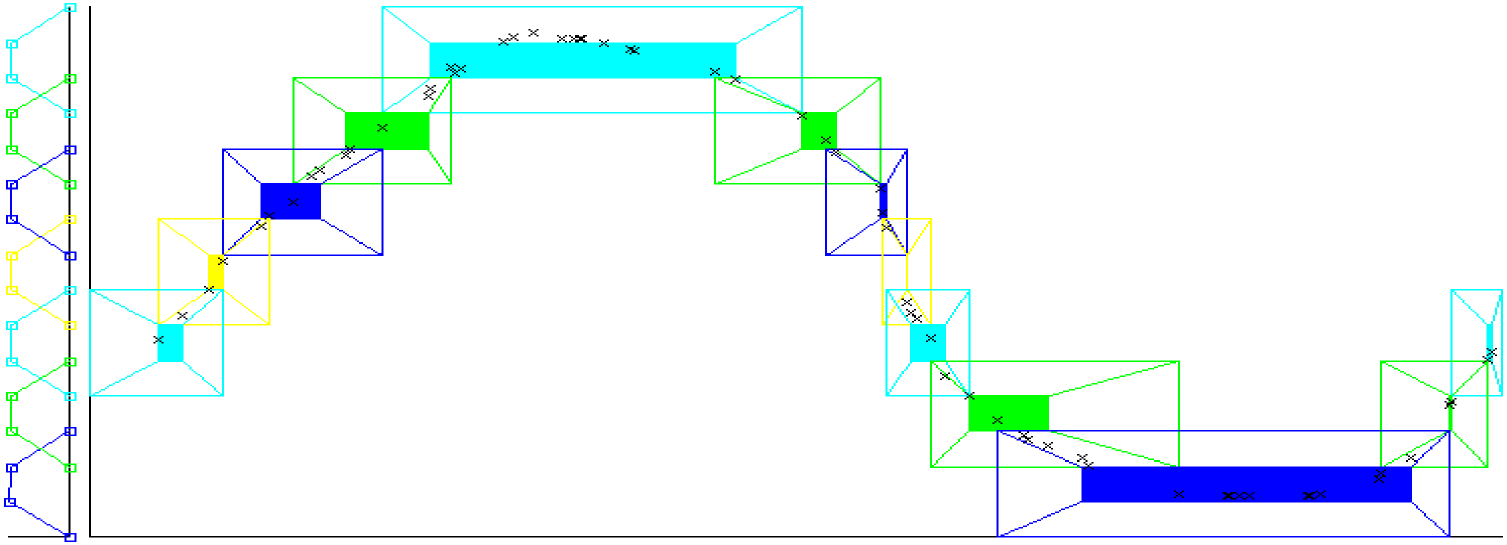
- Projektion von Schnitt mit zylindrischer Erweiterung:

$$B = \mathrm{proj}_y \{(\mathbf{A} \times I) \cap \tilde{f}\}$$



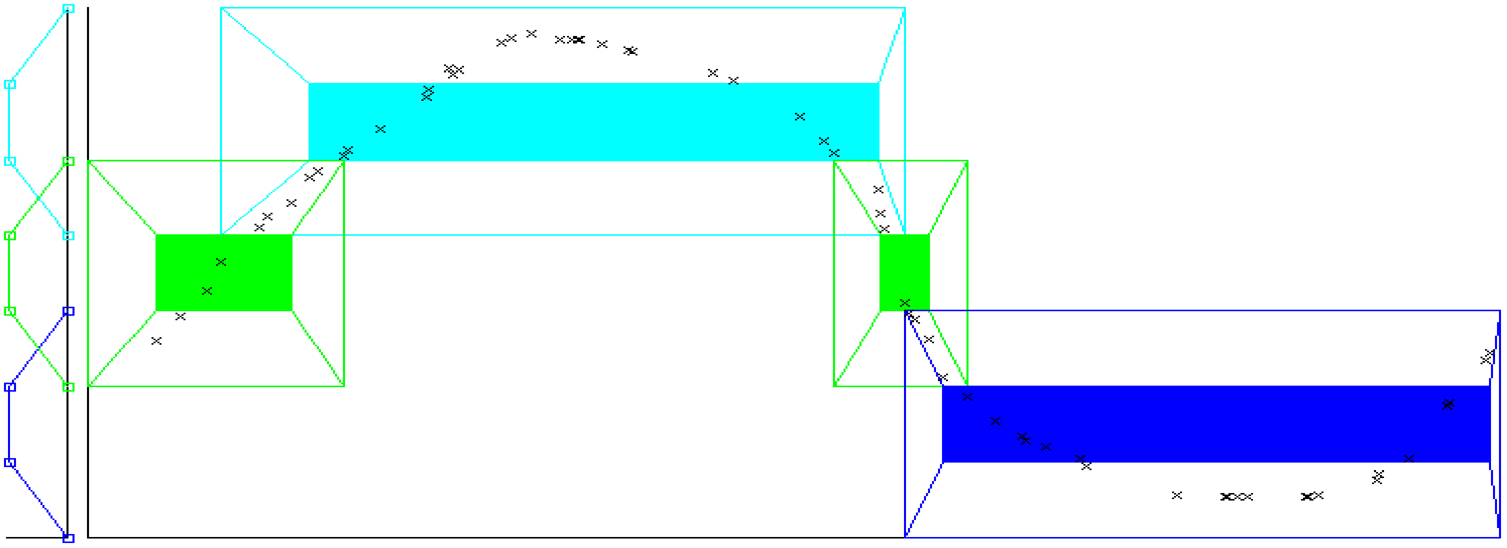
GRANULARITÄT VON FUZZY-GRAPHEN

- Fein granular



GRANULARITÄT VON FUZZY-GRAPHEN

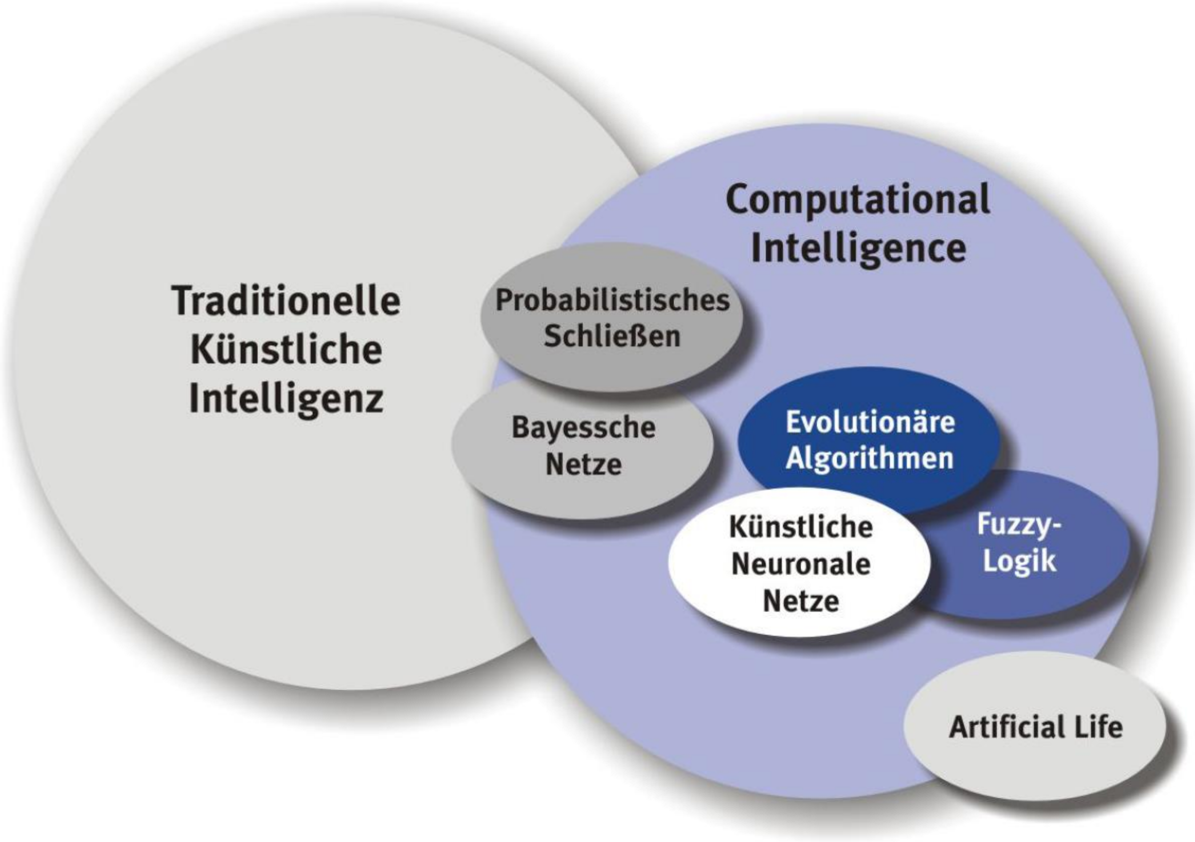
- Grob granular



ZUSAMMENFASSUNG

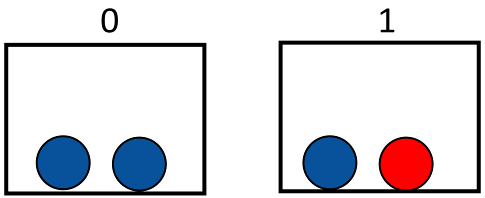
- Fuzzy-Methoden
  - Modellierung unsicheren Wissens (Fuzzy-Regeln)
  - Inferenz berücksichtigt Ungenauigkeiten
- Wesentliche Konzepte:
  - Fuzzy-Mengen und Grad der Mitgliedschaft
  - Operatoren auf Fuzzy-Mengen
  - Erweiterungsprinzip: Anwendung klassischer Funktionen auf Fuzzy-Zahlen
- Algorithmen (meist heuristisch) für:
  - Erzeugung von Fuzzy-Regeln aus Daten
  - Lernen von Fuzzy-Approximatoren (Fuzzy-Graphen/Regelsysteme)
- Fuzzy-Regler haben weite Anwendung gefunden
  - Reiskocher
  - Kameras
  - Waschmaschinen
  - Automatik-Getriebe

BAUSTEINE DER COMPUTATIONAL INTELLIGENCE

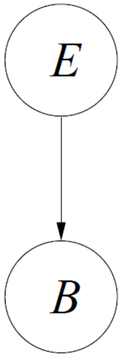


PROBABILISTISCHE GRAFISCHE MODELLE

Beispiel für das Rechnen mit Wahrscheinlichkeiten



- Gegeben:
  - Zwei Eimer mit je zwei Bällen rot/blau
  - 100€ Belohnung bei Ziehen des einzigen roten Balls
  - Zufällige Wahl des Eimers
  - Zufälliges Ziehen eines Balls => Dieser ist blau
- Jetzt darf man den Eimer wechseln und noch einmal ziehen.
  Ist das Wechseln vorteilhaft?

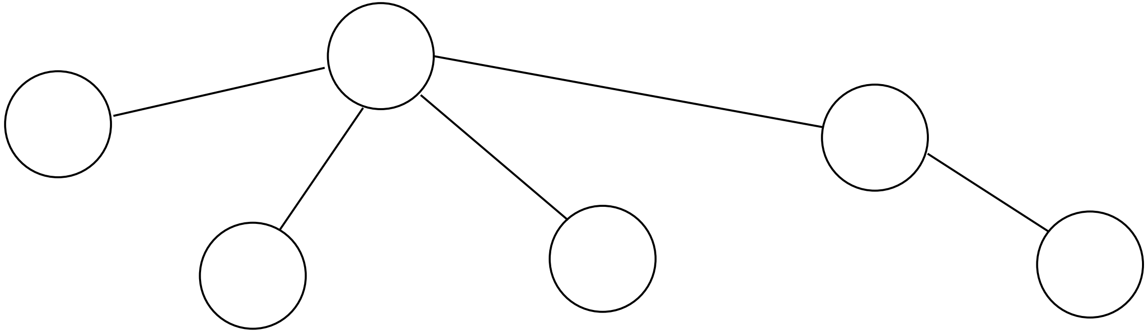


- Zufallsvariablen
  - Gewählter Eimer: $E \in \{1, 0\}$
  - Farbe des Balls: $B \in \{r, b\}$
  - Zufällige Eimerwahl: $P(E = 1) = P(E = 0) = 1/2$
  - Ballverteilung: $P(B = r \mid E = 1) = 1/2$ $P(B = r \mid E = 0) = 0$
  - Frage: $P(E = 1 \mid B = b) \geq 1/2$? // Falls ja, sollten wir nicht wechseln.

$$P(E = 1 \mid B = b) = \frac{P(B{=}b|E{=}1)P(E{=}1)}{P(B{=}b)} = \frac{1/2 \cdot 1/2}{3/4} = 1/3$$ **=> Wechseln!**

MOTIVATION FÜR PROBABILISTISCHE GRAFISCHE MODELLE

- Kompakte Beschreibung von Abhängigkeiten zwischen Zufallsvariablen
- Informationsverarbeitung als Inferenz z.B. durch Nachrichtenaustausch zwischen benachbarten Knoten
- Mögliche Abstraktion neuronaler Informationsverarbeitung
- Viele Algorithmen des maschinellen Lernens können aus grafischen Modellen abgeleitet werden



ERINNERUNG: WAHRSCHEINLICHKEITSRECHNUNG

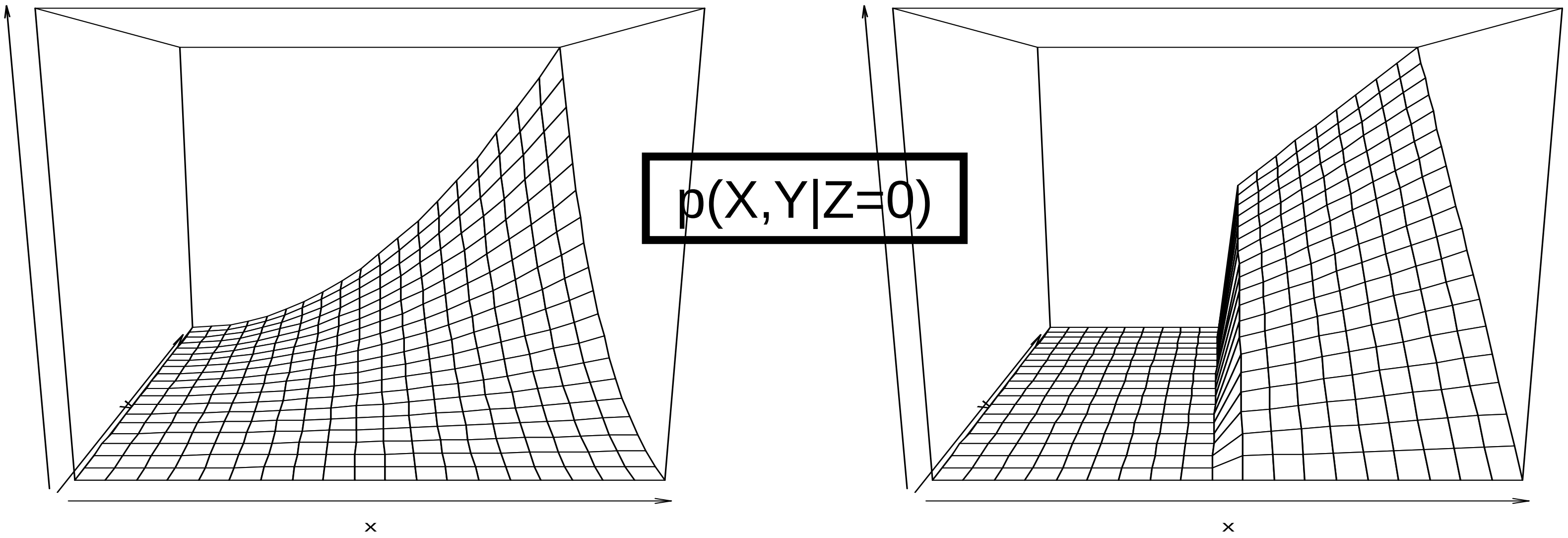
- Zufallsvariablen: X, Y, Z
  - Diskret: Verteilungen beschrieben durch Wahrscheinlichkeiten p(X=Wert)
  - Kontinuierlich: Verteilungen beschrieben durch Wahrscheinlichkeitsdichte p(x)
- Gemeinsame Verteilung: $p(X,Y)$
- Bedingte Verteilung: $$p(X \mid Y) = \frac{p(X,Y)}{p(Y)}$$
- Produktregel: $$p(X,Y) = p(X \mid Y)\,p(Y)$$
- Aussummieren: $$p(x) = \sum_y p(x,y) \qquad p(x) = \int_{-\infty}^{\infty} p(x,y)\,dy$$

UNABHÄNGIGKEIT VON ZUFALLSVARIABLEN

- Unabhängigkeit:
  - X, Y unabhängig g.d.w.: $p(X,Y) = p(X)p(Y)$
  - Mit Produktregel $p(X,Y) = p(X \mid Y)p(Y)$
    - X, Y unabhängig g.d.w.: $p(X \mid Y) = p(X)$
    - X, Y unabhängig g.d.w.: $p(Y \mid X) = p(Y)$
- Bedingte Unabhängigkeit:
  - Anwendung der Unabhängigkeit auf $\mathrm{p}(\mathrm{X},\mathrm{Y}|\mathrm{Z})$
  - X, Y unabhängig gegeben Z g.d.w. $p(X,Y \mid Z) = p(X \mid Z)p(Y \mid Z)$
  - X, Y unabhängig gegeben Z g.d.w. $p(X \mid Y,Z) = p(X \mid Z)$
  - X, Y unabhängig gegeben Z g.d.w. $p(Y \mid X,Z) = p(Y \mid Z)$

BEDINGTE UNABHÄNGIGKEIT

- Interpretation:
  - Sobald Z bekannt ist, sind die Variablen X und Y unabhängig.
  - Bei Kenntnis von Z liefert X keinerlei zusätzliche Information über Y, und umgekehrt.

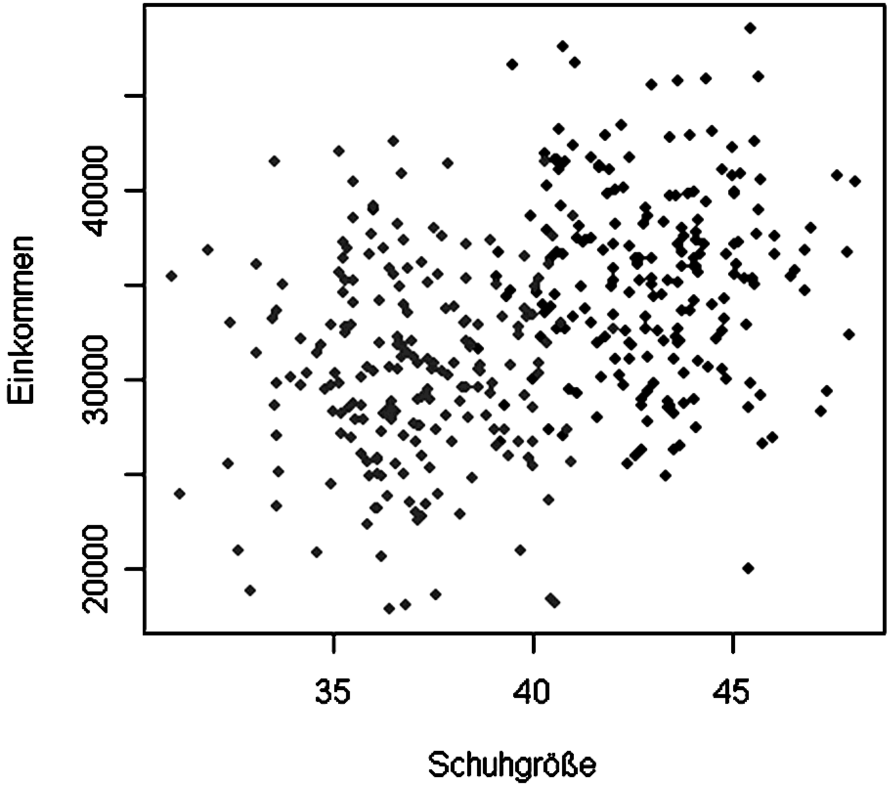


p(X|Z),p(Y|Z) unabhängig

p(X|Z),p(Y|Z) abhängig

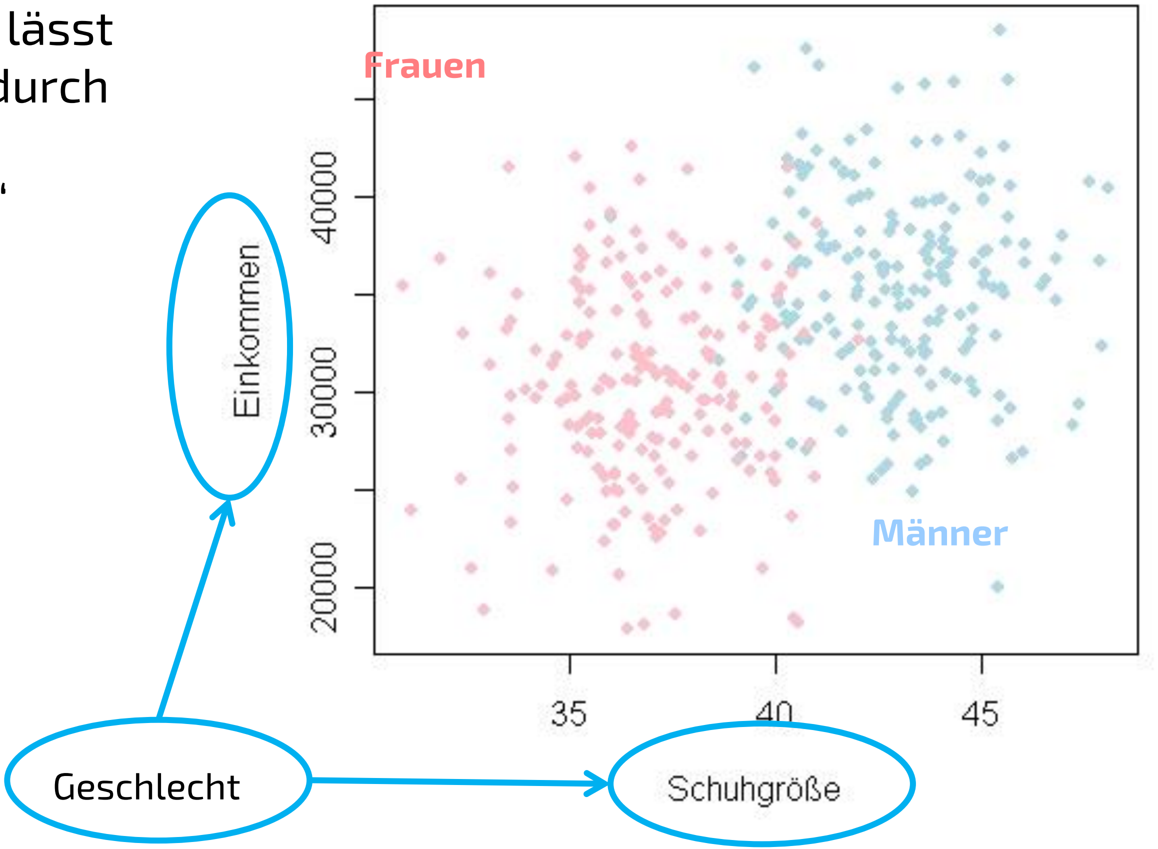
BEISPIEL FÜR BEDINGTE UNABHÄNGIGKEIT

- Das Jahreseinkommen hängt von der Schuhgröße ab.



BEISPIEL FÜR BEDINGTE UNABHÄNGIGKEIT

- Das Jahreseinkommen hängt von der Schuhgröße ab.
- Unterschied lässt sich jedoch durch die Variable „Geschlecht“ erklären.



BAYES-NETZE

- Ein Bayes-Netz ***B*** besteht aus
  - Einem **gerichteten, azyklischen Graphen *G=(V,E)*,** wobei die Knoten ***V={V₁,…,Vₙ}*** Zufallsvariablen sind.
  - Einer Menge **lokaler bedingter Wahrscheinlichkeitsverteilungen *P(V_j | pa(V_j)), j=1,…,n,*** wobei ***pa(V_j)*** die Elternknoten von ***V_j*** sind.

- ***B*** codiert eine gemeinsame Wahrscheinlichkeitsverteilung von ***(V₁,…,Vₙ)*** als Produkt der lokaler Wahrscheinlichkeiten

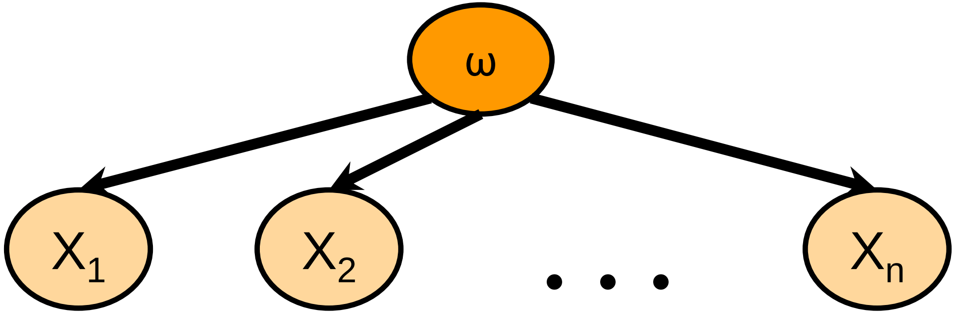
$$\boxed{P(V_1 = v_1, \ldots, V_n = v_n) = \prod_{j=1}^{n} P(V_j = v_j \mid pa(V_j) = v_{pa(V_j)})}$$

BEISPIEL: NAIVE BAYES-ANNAHME

- Naive Bayes-Annahme:

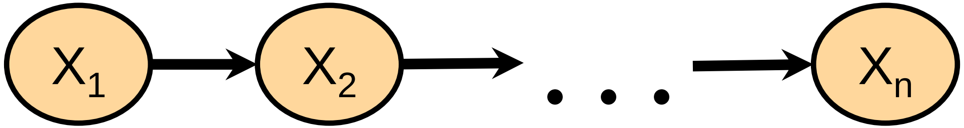
$$P(X_1, \ldots, X_n \mid \omega) = \prod_{k=1}^{n} P(X_k \mid \omega)$$

$$P(X_1, \ldots, X_n, \omega) = P(X_1, \ldots, X_n \mid \omega)\, P(\omega) = P(\omega) \prod_{k=1}^{n} P(X_k \mid \omega)$$



BEISPIEL: MARKOV-KETTE

- Lineare Folge von Knoten:



$$P(X_1,\ldots,X_n) = \prod_{j=1}^{n} P(X_j \mid pa(X_j)) = \mathrm{P}(\mathrm{X}_1)\cdot\prod_{j=2}^{n} P(X_j \mid X_{j-1})$$

- Interpretation: Die Variable X_j hängt nur von ihrer direkten Vorgängervariablen ab
- Markovketten werden oft zur Modellierung von Zustandsfolgen bzw. Zeitreihen $X_1,\ldots,X_n$ verwendet (j ist dann ein Zeitindex).

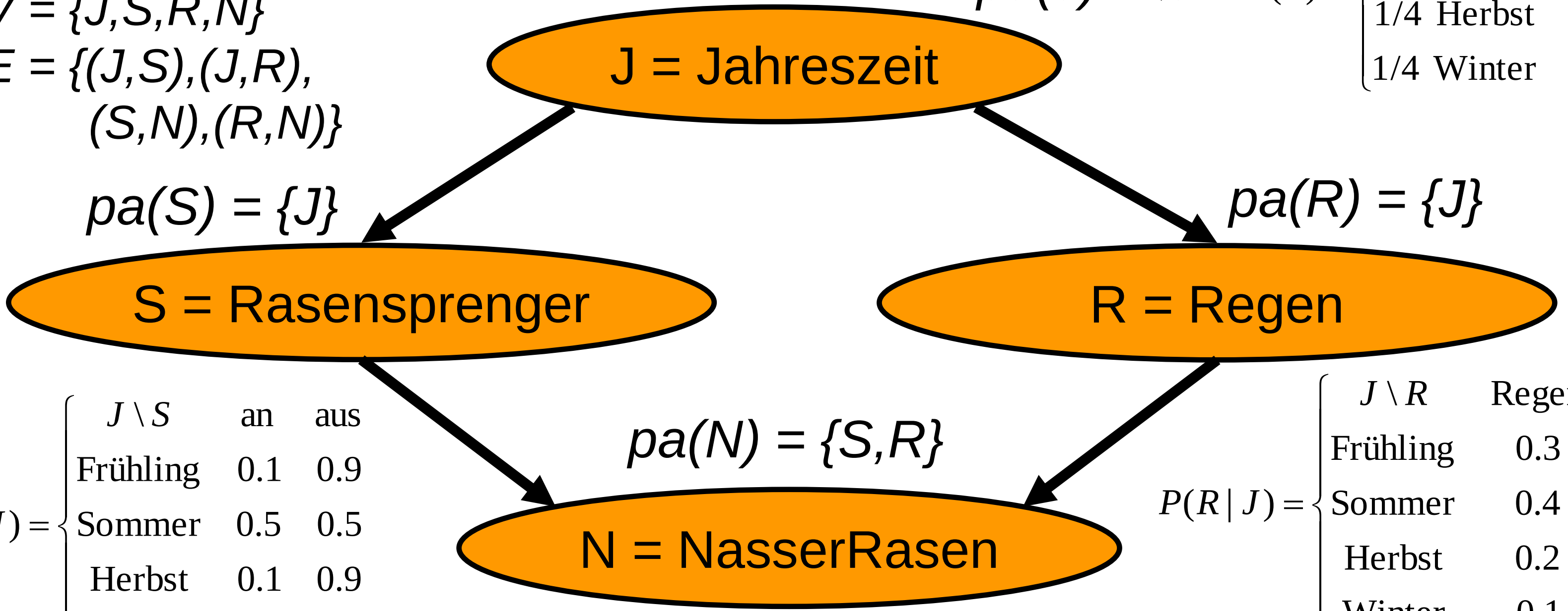
BEISPIEL: SPRINKLER-NETZWERK

$V = \{J,S,R,N\}$

$E = \{(J,S),(J,R),(S,N),(R,N)\}$

$pa(J) = \varnothing$

$$P(J) = \begin{cases} 1/4 & \text{Frühling} \\ 1/4 & \text{Sommer} \\ 1/4 & \text{Herbst} \\ 1/4 & \text{Winter} \end{cases}$$



$$P(S \mid J) = \begin{cases} J \setminus S & \text{an} & \text{aus} \\ \text{Frühling} & 0.1 & 0.9 \\ \text{Sommer} & 0.5 & 0.5 \\ \text{Herbst} & 0.1 & 0.9 \\ \text{Winter} & 0 & 1 \end{cases}$$

$$P(R \mid J) = \begin{cases} J \setminus R & \text{Regen} & \text{kein Regen} \\ \text{Frühling} & 0.3 & 0.7 \\ \text{Sommer} & 0.4 & 0.6 \\ \text{Herbst} & 0.2 & 0.8 \\ \text{Winter} & 0.1 & 0.9 \end{cases}$$

$$P(N \mid S,R) = \begin{cases} (S,R) \setminus N & \text{nass} & \text{trocken} \\ (\text{an, Regen}) & 1 & 0 \\ (\text{an, kein Regen}) & 1 & 0 \\ (\text{aus, Regen}) & 0.9 & 0.1 \\ (\text{aus, kein Regen}) & 0.01 & 0.99 \end{cases}$$

SPRINKLER-NETZWERK II

- Die gemeinsame Verteilung von {J,S,R,N} wird durch den Graphen und 3+4+4+4 =15 reelle Zahlen codiert:

$$pa(.) = \begin{cases} J \mapsto \{\ \} \\ S \mapsto \{J\} \\ R \mapsto \{J\} \\ N \mapsto \{S,R\} \end{cases}$$

$$P(J) = \begin{cases} \boxed{1/4} & \text{Frühling} \\ \boxed{1/4} & \text{Sommer} \\ \boxed{1/4} & \text{Herbst} \\ 1/4 & \text{Winter} \end{cases}$$

$$P(S \mid J) = \begin{cases} J \setminus S & \text{an} & \text{aus} \\ \text{Frühling} & \boxed{0.1} & 0.9 \\ \text{Sommer} & \boxed{0.5} & 0.5 \\ \text{Herbst} & \boxed{0.1} & 0.9 \\ \text{Winter} & \boxed{0} & 1 \end{cases}$$

$$P(R \mid J) = \begin{cases} J \setminus R & \text{Regen} & \text{kein Regen} \\ \text{Frühling} & \boxed{0.3} & 0.7 \\ \text{Sommer} & \boxed{0.4} & 0.6 \\ \text{Herbst} & \boxed{0.2} & 0.8 \\ \text{Winter} & \boxed{0.1} & 0.9 \end{cases}$$

$$P(N \mid S,R) = \begin{cases} (S,R) \setminus N & \text{nass} & \text{trocken} \\ (\text{an, Regen}) & \boxed{1} & 0 \\ (\text{an, kein Regen}) & \boxed{1} & 0 \\ (\text{aus, Regen}) & \boxed{0.9} & 0.1 \\ (\text{aus, kein Regen}) & \boxed{0.01} & 0.99 \end{cases}$$

$$\boxed{\begin{aligned} P(J=j, S=s, R=r, N=n) = \\ P(N=n \mid S=s, R=r) \cdot P(S=s \mid J=j) \cdot P(R=r \mid J=j) \cdot P(J=j) \end{aligned}}$$

z.B.

$$\begin{aligned} & P(J=\text{Sommer}, S=\text{aus}, R=\text{Regen}, N=\text{nass}) \\ &= P(N=\text{nass} \mid S=\text{aus}, R=\text{Regen}) \cdot P(R=\text{Regen} \mid J=\text{Sommer}) \\ &\quad \cdot P(S=\text{aus} \mid J=\text{Sommer}) \cdot P(J=\text{Sommer}) \\ &= 0.9 \cdot 0.4 \cdot 0.5 \cdot 0.25 \\ &= 0.045 \end{aligned}$$

KOMPAKTE DARSTELLUNG

- Die vollständige Festlegung der gemeinsamen Verteilung von $\{J,S,R,N\}$ in Form einer Wahrscheinlichkeitstabelle hätte $4 \cdot 2 \cdot 2 \cdot 2 - 1 = 31$ reelle Zahlen benötigt.
  => Exponentiell kompaktere Darstellung

- Notwendige Folge:
  Nicht alle Wahrscheinlichkeitsverteilungen können mit diesem Bayes-Netz exakt codiert werden

- Aber: Jede Verteilung kann als Bayes-Netz dargestellt werden, es ist nur nicht klar, dass dieses klein ist.

- Ergo: Es gibt geschickte und ungeschickte Darstellungen als Bayes-Netz

BEISPIEL: ALARM-NETZWERK

- Modelliertes Szenario
  - Unser Haus in LA hat eine Alarmanlage.
  - Wir sind im Urlaub. Unser Nachbar ruft an, falls er den Alarm hört. Wenn eingebrochen wurde, wollen wir zurück kommen.
  - Leider ist der Nachbar nicht immer zu Hause.
  - Leider geht die Alarmanlage auch bei kleinen Erdbeben los.
- Fünf binäre Zufallsvariablen
  - B Burglary – Einbruch hat stattgefunden
  - E Earthquake – Erdbeben hat stattgefunden
  - A Alarm – Alarmanlage geht los
  - N NeighborCalls – Nachbar ruft an
  - R RadioReport – Bericht über Erdbeben im Radio

ALARM: BEDINGTE UNABHÄNGIGKEITEN

- Zerlegung in Faktoren nach Produktregel

$$p(B,E,A,N,R) = p(B)\,p(E \mid B)\,p(A \mid B,E)\,p(N \mid B,E,A)\,p(R \mid B,E,A,N)$$

  - Geht immer, verschiedene Reihenfolgen möglich => Wir haben noch nichts gewonnen.

- Annahme bedingter Unabhängigkeiten

$p(E \mid B) = p(E)$	*Erdbeben hängt nicht von Einbruch ab*
$p(A \mid B,E) = p(A \mid B,E)$	*Alarm hängt von Einbruch und Erdbeben ab*
$p(N \mid B,E,A) = p(N \mid A)$	*Anruf von Nachbar hängt nur von Alarm ab*
$p(R \mid B,E,A,N) = p(R \mid E)$	*Nachricht im Radio hängt nur Erdbeben ab*

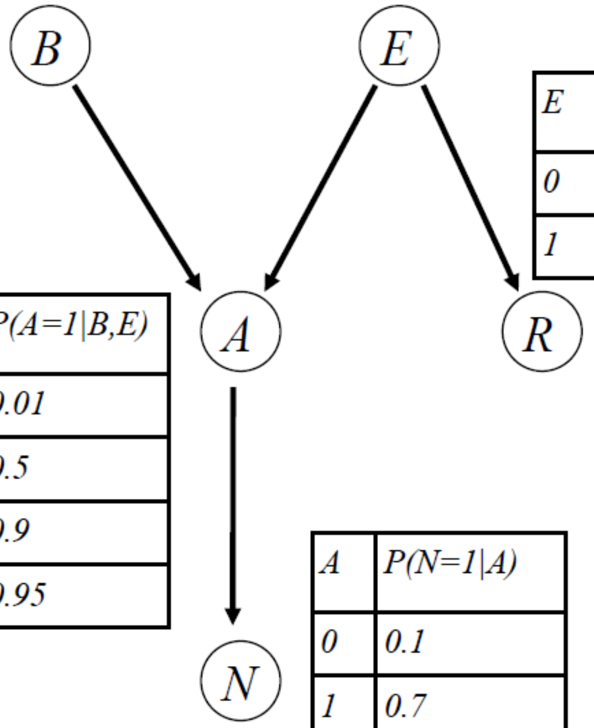
  - Gewählte Reihenfolge ermöglicht es, diese Annahmen zu machen.

- Vereinfachte Darstellung der Verbundverteilung

$$p(B,E,A,N,R) = p(B)\,p(E)\,p(A \mid E,B)\,p(N \mid A)\,p(R \mid E)$$

ALARM-NETZWERK: MODELLIERTE VERTEILUNG

$$p(B,E,A,N,R) = p(B)p(E)p(A \mid E,B)p(N \mid A)p(R \mid E)$$



$P(B=1)$
0.1

$P(E=1)$
0.2

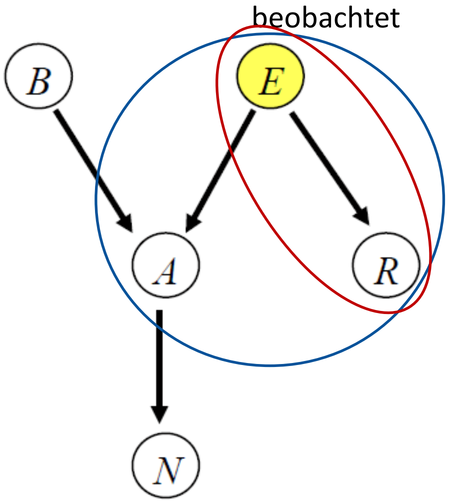
E	$P(R=1 \mid E)$
0	0.01
1	0.5

B	E	$P(A=1 \mid B,E)$
0	0	0.01
0	1	0.5
1	0	0.9
1	1	0.95

A	$P(N=1 \mid A)$
0	0.1
1	0.7

UNABHÄNGIGKEIT: D-SEPARATION

- Menge einfacher Regeln, nach denen sich alle Unabhängigkeiten ableiten lassen
- „D“ steht für Directed, weil gerichteter Graph



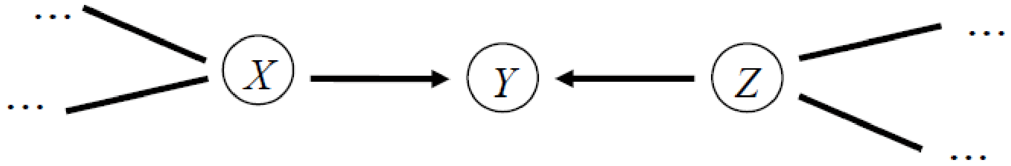
$$p(B,E,A,N,R) =$$
$$p(B)p(E)p(A \mid E,B)p(N \mid A)p(R \mid E)$$

- Betrachte Pfad $A \leftarrow E \rightarrow R$
- Gilt $A \perp R \mid E$ (Unabhängigkeit von A und R gegeben E)?
- Ja, $p(A \mid R,E) = p(A \mid E)$

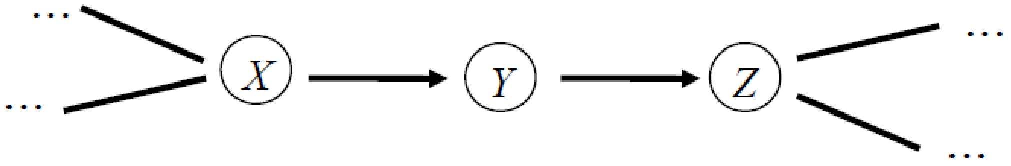
  $p(A|R,E) = p(A,R,E)/p(R,E) = p(A|E)\cancel{p(R|E)p(E)}/\cancel{p(R|E)p(E)}$
- Intuitiv:
  - Wenn wir schon wissen, dass ein Erdbeben eingetreten ist, wird die Wahrscheinlichkeit für Alarm nicht höher/niedriger durch RadioReport
  - Der divergierende Pfad $A \leftarrow E \rightarrow R$ wird durch Beobachtung von E blockiert

VERBINDUNGSMUSTER

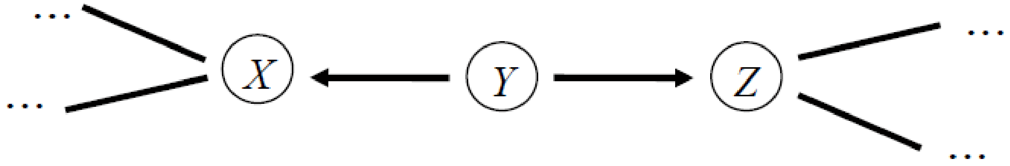
- Konvergenz bei Y



- Serielle Verbindung bei Y

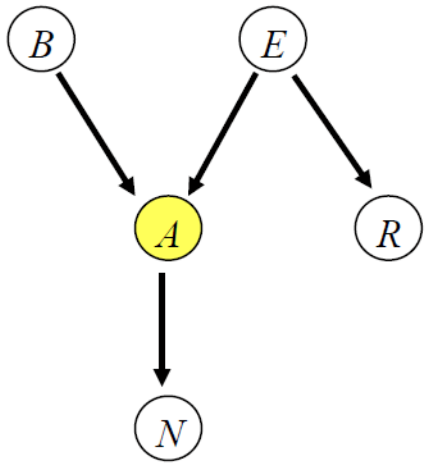


- Divergenz bei Y



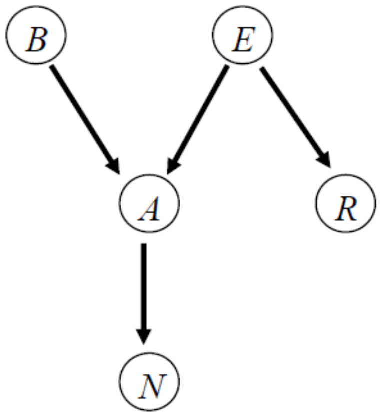
D-SEPARATION

- Beispiele



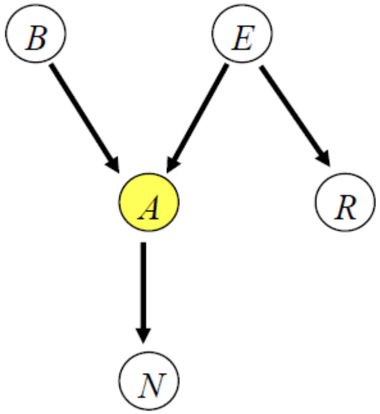
Gilt $B \perp N \mid A$?

Ja.



Gilt $B \perp E \mid \varnothing$?

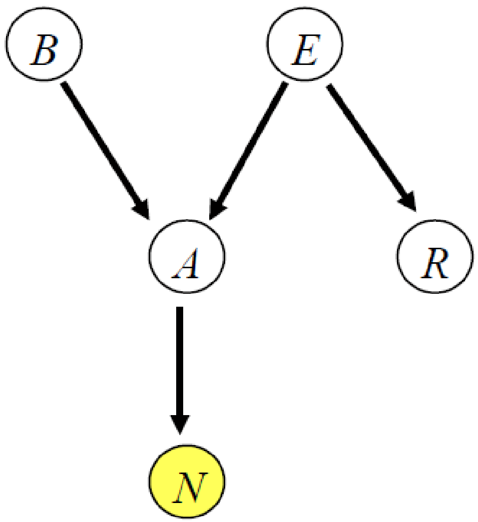
Ja.



Gilt $B \perp E \mid A$?

Nein!

D-SEPARATION



Gilt $B \perp E \mid N$?

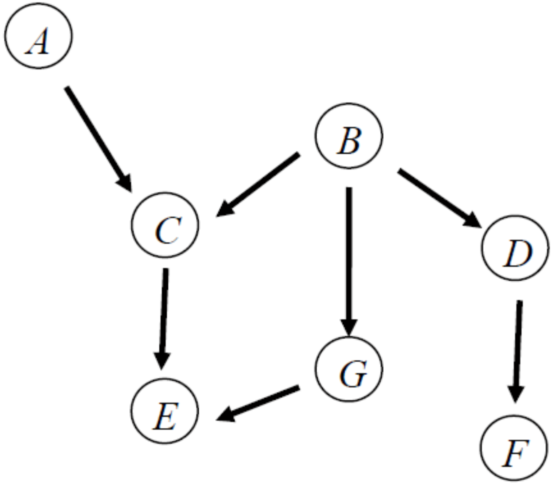
Nein.

- Ein Pfad …*X*-*Y*-*Z*… ist blockiert bei *Y*, wenn
  - Divergierende Verbindung und *Y* beobachtet, oder
  - Serielle Verbindung und *Y* beobachtet, oder
  - Konvergierende Verbindung und weder *Y* noch einer seiner Nachfahren beobachtet
  - Sonst ist der Pfad frei bei *Y*

- Ein Pfad ist insgesamt blockiert, wenn er bei einem auf dem Pfad liegenden Knoten blockiert ist

D-SEPARATION

- Seien $\mathbb{X}$, $\mathbb{Y}$, $\mathbb{Z}$ disjunkte Mengen von Zufallsvariablen.
- Definition: $\mathbb{X}$ und $\mathbb{Y}$ sind D-separiert durch $\mathbb{Z}$ g.d.w jeder Pfad von einem Knoten $X \in \mathbb{X}$ zu einem Knoten $Y \in \mathbb{Y}$ blockiert ist gegeben $\mathbb{Z}$.
- Dann: $p(\mathbb{X}|\mathbb{Y},\mathbb{Z}) = p(\mathbb{X}|\mathbb{Z})$ und $p(\mathbb{Y}|\mathbb{X},\mathbb{Z}) = p(\mathbb{Y}|\mathbb{Z})$



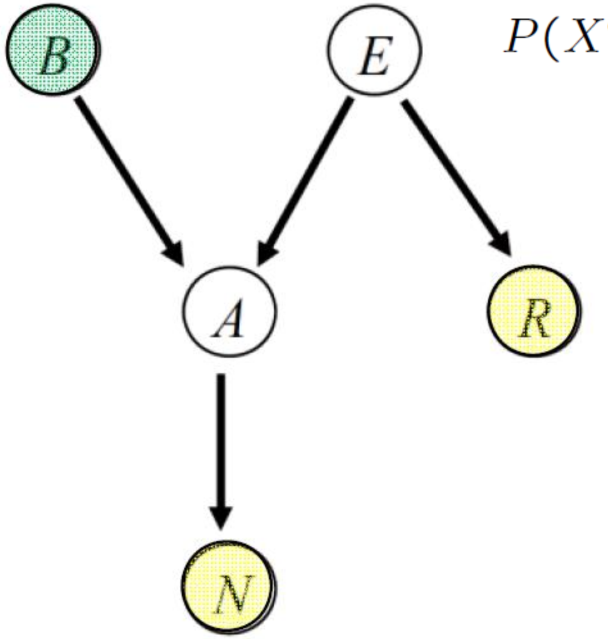
Gilt $A \perp F | D$? Ja.

Gilt $B \perp E | C$? Nein: B -> G -> E

Gilt $A \perp E | C$? Nein: A->C->B->G->E

INFERENZ

- Gegeben: Beobachte Variablen $\mathcal{X}^m$ (Evidenz $\{N, R\}$)
- Berechne Randverteilung über Anfrage-Variablen X^q ($\{B\}$), gegeben Evidenz
- Allgemeine Methode: Marginalisierung über restliche Variablen $\mathcal{X}^r$ ($\{A, E\}$)

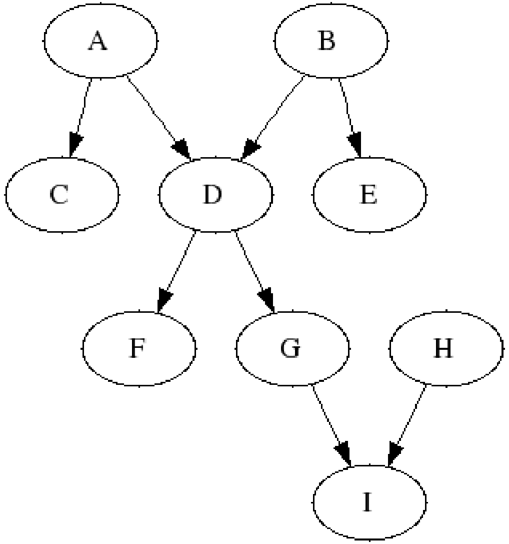


$$P(X^q, \mathcal{X}^m) = \sum_{\mathcal{X}^r} P(X_1, \ldots X_M) \; ; \; P(\mathcal{X}^m) = \sum_{\mathcal{X}^q} P(X^q, \mathcal{X}^m)$$

$$P(X^q | \mathcal{X}^m) = \frac{P(X^q, \mathcal{X}^m)}{P(\mathcal{X}^m)}$$

- Wahrscheinlichkeit für Einbruch (B=1) gegeben dass der Nachbar uns angerufen hat (N=1)?
  - Kein Radiobericht über Erdbeben: $p(B=1|N=1, R=0)=0.6$
  - Radiobericht über Erdbeben: $p(B=1|N=1, R=1)=0.2$

INFERENZ IN AZYKLISCHEN BAYES'SCHEN NETZEN



- Marginalisierung teuer
  - Summe hat exponentiell [in der Anzahl der unbeobachteten Variablen] viele Terme
- Nutze Struktur des Netzes für effiziente Marginalisierung!

BELIEF PROPAGATION

- Inferenz durch Nachrichtenaustausch zwischen benachbarten Knoten
  - Nachrichten von Eltern zu Kindern

$$\pi_{Y_j}(x) = \left[\prod_{k \neq j} \lambda_{Y_k}(x)\right] \sum_{u_1,\ldots,u_n} P(x|u_1,\ldots,u_n) \prod_i \pi_X(u_i)$$

  - Nachrichten von Kindern zu Eltern

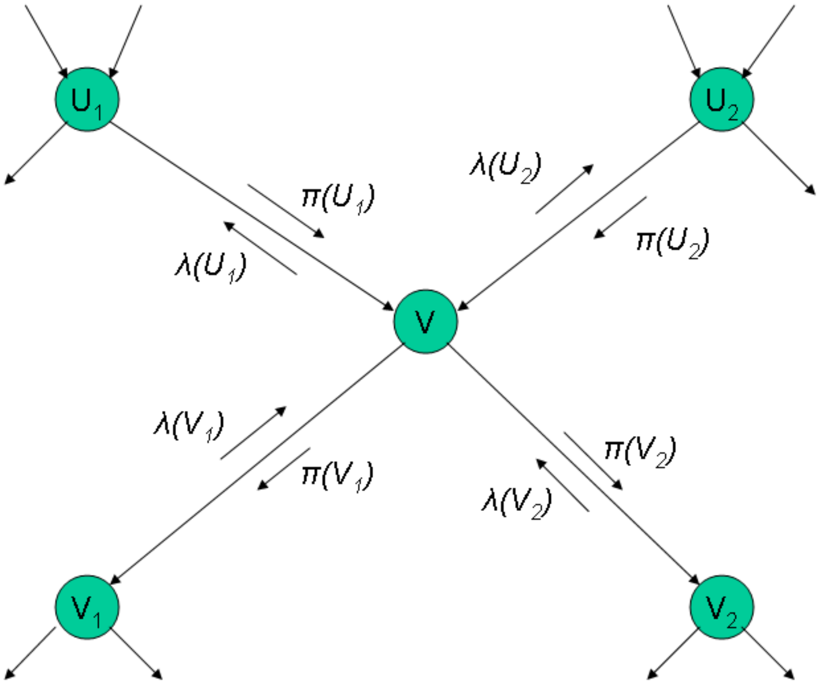
$$\lambda_X(u_i) = \sum_x \left[\lambda(x) \sum_{u_k:k\neq i} P(x|u_1,\ldots,u_n) \prod_{k\neq i} \pi_X(u_k)\right]$$

$$\lambda(x) = \prod_j \lambda_{Y_j}(x)$$

  - Berechnung des Beliefs

$$\pi(x) = \sum_{u_1,\ldots,u_n} P(x|u_1,\ldots,u_n) \prod_i \pi_X(u_i)$$

$$Bel(x) = P(x|Evidence) = \frac{1}{Z}\lambda(x)\pi(x)$$



EXAKTE VS. APPROXIMATIVE INFERENZ

- Exakte Inferenz
  - Inferenz schwieriges Problem
  - Allgemeine Bayes‘sche Netze: Exakte Inferenz NP-hart
  - Es gibt Algorithmen für exakte Inferenz in allgemeinen Bayes´schen Netzen, deren Laufzeit von den Eigenschaften der Graphstruktur abhängt (Message-Passing)
- Techniken für approximative Inferenz
  - Sampling
  - Variationale Inferenz
  - Expectation Propagation

ZUSAMMENFASSUNG GRAPHISCHE MODELLE

- Modellierung der Verbundverteilung einer Menge von Zufallsvariablen $\{X_1, \ldots, X_N\}$

$$p(X_1, \ldots, X_N) = \prod_{i=1}^{N} p(X_i \mid pa(X_i))$$

  - Gerichteter azyklischer Graph drückt Unabhängigkeitsannahmen aus
  - Kompakt, intuitiv verständlich
  - Bedingte Unabhängigkeiten zwischen Variablenmengen können an der Graphstruktur abgelesen werden (D-Separation)
  - Effiziente Inferenzalgorithmen nutzen Graphstruktur aus

KLAUSUR

- 20. Juli ~~15:00-17:00 im HS 1+2 (Hörsaalzentrum, Endenicher Allee 19c)~~

  Geändert: 12:00-14:00 im Wolfgang Paul-Hörsaal, Kreuzbergweg 28
- 90 Minuten Bearbeitungszeit
- Closed Book
- Testklausur ist im Verzeichnis http://www.ais.uni-bonn.de/SS20/CI