Machine Intelligence 1

3.1 Uncertainty and Inference

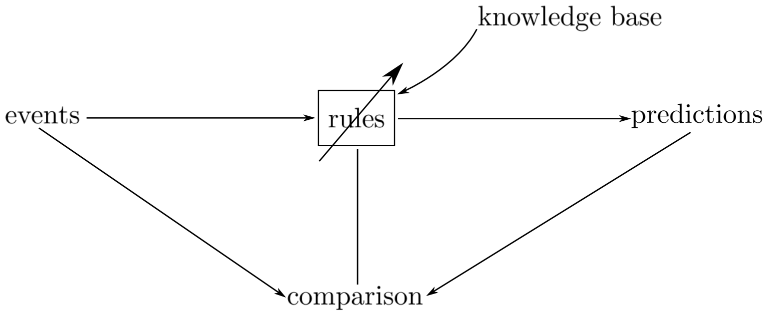
Prof. Dr. Klaus Obermayer

Fachgebiet Neuronale Informationsverarbeitung (NI)

WS 2017/2018

3.1.1 Degrees of Belief

Reasoning and Learning



What about logic?

1. $\forall p : \text{symptom}(p, toothache) \Rightarrow \text{disease}(p, cavity)$
   - rule is wrong: not all patients with toothache have cavities
2. $\forall p : \text{symptom}(p, toothache) \Rightarrow$ $\text{disease}(p, cavity) \vee \text{disease}(p, gumdisease) \vee \ldots$
   - almost unlimited list of possible causes
3. $\forall p : \text{disease}(p, cavity) \Rightarrow \text{symptom}(p, toothache)$
   - rule is wrong: not all cavities cause pain

- How can we make decisions when we are never 100% sure?

Degrees of belief (1)

$$P(H) : H \to [0, 1] \qquad \text{assignment of numbers}$$

$$P(H) = 0 \qquad H \text{ is false}$$

$$P(H) = 1 \qquad H \text{ is true}$$

$$0 < P(H) < 1 \qquad \text{quantifies our degree of belief}$$

- $P(H)$ obeys the laws of probability theory
- but: no justification via repeated observations

E. T. Jaynes: *Probability theory – the logic of science* (2003)

Degrees of belief (2)

Application to betting agents (de Finetti, 1931)

"If Agent 1 expresses a set of degrees of belief that violate the axioms of probability theory, then there is a combination of fair bets by Agent 2 that guarantees that Agent 1 will loose money all the time."

- Agent 1 has inconsistent beliefs
  - violates $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Agent 1		Agent 2		Outcome for Agent 1			
Proposition	Belief	Bet	Stakes	$A \wedge B$	$A \wedge \neg B$	$\neg A \wedge B$	$\neg A \wedge \neg B$
A	0.4	A	4 to 6	-6	-6	4	4
B	0.3	B	3 to 7	-7	3	-7	3
$A \vee B$	0.8	$\neg(A \vee B)$	2 to 8	2	2	2	-8
				-11	-1	-1	-1

- Agent 2 can devise a fair bet which always wins, no matter the outcome of A and B.

(see blackboard)

3.1.2 The Description of the World

Random variables and their domains

- Boolean variables, e.g. $cavity \in \{\text{true}, \text{false}\}$
  - proposition: $cavity = \text{true}$
- discrete ordinal variables, e.g. $weather \in \{\text{sunny}, \text{rainy}, \text{cloudy}\}$
  - proposition: $weather = \text{sunny}$
- continuous variables, e.g. $temperature \in \mathbb{R}_0^+$
  - proposition: $temperature \in [290K, 291K]$
- description of the world: complete set of all variables

Atomic events

- atomic event: one assignment of all random variables
  - atomic events are mutually exclusive
  - set of atomic events must be exhaustive

- proposition: disjunction of atomic events
  - e.g. $cavity = \text{true}$ is equivalent to:
    $(cavity = \text{true} \wedge toothache = \text{false} \wedge catch = \text{false}) \vee$
    $(cavity = \text{true} \wedge toothache = \text{true} \wedge catch = \text{true}) \vee \ldots$

Prior (unconditional) probabilities

- probabilities of atomic events
  - e.g. $P(cavity = \text{true} \wedge toothache = \text{true} \wedge catch = \text{true}) = 0.108$
- domain knowledge:
  set of unconditional probabilities for all atomic events

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

$P(toothache, cavity, catch)$

Conditional probabilities

- $P(X|e)$: degree of belief in X, given *all* we know is evidence e
- e.g. $P(cavity = \text{true} \mid toothache = \text{true}) = 0.8$
  $$P(\underbrace{cavity = \text{false}}_{\text{proposition}} \mid \underbrace{toothache = \text{true}}_{\substack{\text{observation}\\ \text{or evidence}}}) = 0.2$$
- defined by the product rule: $P(X|e) = \frac{P(X,e)}{P(e)}$

3.1.3 Inference

Computing conditional probabilities

- probability of event x given observations $\underline{\mathbf{e}}$
  - X: query variable
  - E_j: evidence variables
  - Y_j: unobserved variables

$$P(x|\underline{\mathbf{e}}) = \frac{P(x,\underline{\mathbf{e}})}{P(\underline{\mathbf{e}})} = \overbrace{\alpha\, P(x,\underline{\mathbf{e}})}^{\text{normalization}} = \overbrace{\alpha \sum_{\underline{\mathbf{y}}} P(x,\underline{\mathbf{e}},\underline{\mathbf{y}})}^{\text{marginalization}}$$

- normalization α can be computed...
  - ...explicitly by $\frac{1}{\alpha} = P(\underline{\mathbf{e}}) = \sum_{x,\underline{\mathbf{y}}} P(x,\underline{\mathbf{e}},\underline{\mathbf{y}})$
  - ...implicitly by ensuring $\sum_{x} P(x|\underline{\mathbf{e}}) = 1$

Computing conditional probabilities

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

- calculate $P(cavity)$ (see blackboard)

Computing conditional probabilities

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

- calculate $P(cavity)$

$cavity$	$\neg cavity$
0.2	0.8

- calculate $P(cavity \,|\, toothache = \text{true})$ (see blackboard)

Computing conditional probabilities

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.144	0.576

- calculate $P(cavity)$

$cavity$	$\neg cavity$
0.2	0.8

- calculate $P(cavity \,|\, toothache = \text{true})$

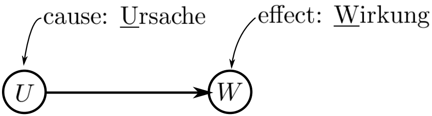
$cavity$	$\neg cavity$
0.6	0.4

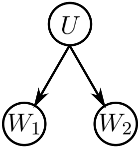
Computational complexity

- Inference requires complete table of prior (unconditional) probabilities.
- N binary random variables X_i:
  - table of joint probabilities has 2^N entries
  - summation over approx. 2^N entries for inference
  - $N = 100 \Rightarrow 2^N \approx 1.3 \cdot 10^{30}$
- This procedure does not scale: additional assumptions are needed.

3.1.4 Conditional Independence

Cause and effect





"causal rule" $P(W|U)$

$P(W_1, W_2|U) = P(W_1|U)\, P(W_2|U)$

Conditional independence

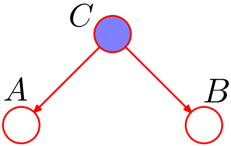
Two random variables X and Y are conditionally independent given Z if:

$$P(X,Y|Z) \quad = \quad P(X|Z) \;\; P(Y|Z) \,.$$

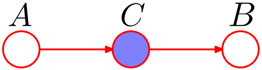
We write $X \perp Y \mid Z$.

Conditional independence examples

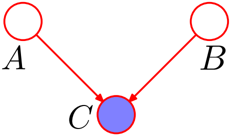
$$X \perp Y \mid Z \qquad \Leftrightarrow \qquad P(X,Y|Z) \quad = \quad P(X|Z)\ P(Y|Z)$$



$A \perp B \mid C$
$A \not\perp B \mid \emptyset$

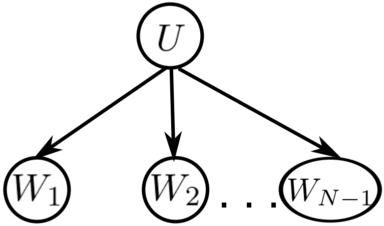


$A \perp B \mid C$
$A \not\perp B \mid \emptyset$



$A \not\perp B \mid C$
$A \perp B \mid \emptyset$

The Naïve Bayes ansatz



- set of binary variables: W_i $(i = 1, \ldots, N-1)$ and U

- assumtion $W_i \perp W_j \,|\, U$, $\forall i, j$

$$\underbrace{P(W_1, W_2, \ldots, W_{N-1}, U)}_{2^N - 1 \text{ independent table entries}} \quad = \quad \underbrace{P(U) \prod_{i=1}^{N-1} P(W_i|u)}_{1+2(N-1) = 2N-1 \text{ independent table entries}}$$

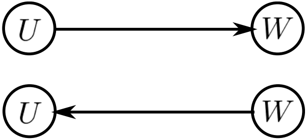
- reduction of computational complexity from $O(2^N)$ to $O(N)$

3.1.5 Bayes' Theorem

Common inference tasks

causal rule: $P(W|U)\,P(U)$

diagnostic rule: $P(U|W)\,P(W)$



(arrows denote statistical dependencies, not causation)

Bayes' theorem

$$P(W|U)\,P(U) \;=\; P(W,U) \;=\; P(U|W)\,P(W) \qquad \text{(product rule)}$$

$$P(U|W) \;=\; \frac{P(W|U)\,P(U)}{P(W)} \;=\; \alpha\,P(W|U)\,P(U) \qquad \text{(Bayes' theorem)}$$

End of Section 3.1

the following slides contain

OPTIONAL MATERIAL

Probabilities of stochastic variables

- for stochastic variables $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ must hold:

$$P(x) \geq 0\,, \quad \forall x \in \mathcal{X}\,, \qquad \sum_{x \in \mathcal{X}} P(x) = 1 \qquad \text{(distribution)}$$

$$P(x) = \sum_{y \in \mathcal{Y}} P(x, y)\,, \qquad \forall x \in \mathcal{X} \qquad \text{(marginalization)}$$

$$P(x, y) = P(x|y)\, P(y)\,, \qquad \forall x \in \mathcal{X}\,, \forall y \in \mathcal{Y} \qquad \text{(product law)}$$

Example domain knowledge

	$toothache$		$\neg toothache$	
	$catch$	$\neg catch$	$catch$	$\neg catch$
$cavity$	0.108	0.012	0.072	0.008
$\neg cavity$	0.016	0.064	0.114	0.576

$$P(toothache, cavity, catch)$$

	$catch$	$\neg catch$
$cavity$	0.54	0.06
$\neg cavity$	0.08	0.32

$$P(cavity, catch \,|\, toothache = \text{true})$$

$cavity$	$\neg cavity$
0.87	0.13

$$P(cavity \,|\, toothache = \text{true}, catch = \text{true})$$