Artificial Intelligence 2

I Artificial Intelligence

II Problem Solving

III Knowledge, Reasoning, Planning

IV Uncertain Knowledge and Reasoning

V Machine Learning

VI Communicating, Perceiving, and Acting

VII Conclusions

Overview

I Artificial Intelligence

II Problem Solving

III Knowledge, Reasoning, Planning

IV Uncertain Knowledge and Reasoning

- **12. Quantifying Uncertainty**
- 13. Probabilistic Reasoning
- 14. Probabilistic Reasoning over Time
- 15. Probabilistic Programming
- 16. Making Simple Decisions
- 17. Making Complex Decisions
- 18. Multiagent Decision Making

V Machine Learning

VI Communicating, Perceiving, and Acting

VII Conclusions

Quantifying Uncertainty

- Acting under Uncertainty
- Basic Probability Notation
- Inference Using Full Joint Distributions
- Independance
- Bayes‘ Rule and Its Use
- Naive Bayes Models
- The Wumpus World Revisited

Acting under Uncertainty

- Agents in the world need to handle under uncertainty
  - Mainly due to partial observability and non-determinism
  - Logical solution: Belief state and contingency plan
    - Suitable only for simple problems
- Example1: Duration and starting time to get by car to the airport?
- Example2: Medical diagnosis
  - Toothache $\Rightarrow$ Cavity (wrong)
  - Toothache $\Rightarrow$ Cavity $\vee$ GumProblem $\vee$ Abscess $\vee$ … (too many potential problems)
  - Cavity $\Rightarrow$ Toothache (better, but still wrong)

Problems with Logical Approach

- **Theoretical ignorance:** Exact rules (sentences) are not known
- **Laziness:** Even it they are known, it is too much work to write exact rules (with complete sets)
- **Practical ignorance:** Even if they are known and written down, we cannot perform all necessary tests in a concrete case

- Alternate approach:
  - **Degree of belief** in relevant sentences
  - Summarizing the uncertainty from laziness and ignorance
  - Foundation: **Probability theory**
  - Depending on the knowledge state (e.g. concerning evidence variables), different probabilities can be assigned to a statement (e.g. for a dental diagnosis like cavity)

Uncertainty and Rational Decisions

- Knowing probabilities of a statement is not sufficient of making decisions
  - Example: Consider a plan A_{90} for getting to the airport in 90 minutes with a 97% chance of catching the flight. A plan A_{180} would have a higher chance, but what plan should be prefered?
- Additional knowledge necessary: Preferences (**utilities**) about the outcome of various plans
  - Example: Utilities of catching the flight and waiting time at the airport

- **Decision Theory = Probability Theory + Utility Theory**
- **Principle of Maximum Expected Utility (MEU):**
  - An Agent ist rational if and only if it chooses the action that yields the highest expected utility, averaged over all the possible outcomes of the action

Decision-theoretic Agent

```
function DT-AGENT(percept) returns an action
  persistent: belief_state, probabilistic beliefs about the current state of the world
              action, the agent's action

  update belief_state based on action and percept
  calculate outcome probabilities for actions,
      given action descriptions and current belief_state
  select action with highest expected utility
      given probabilities of outcomes and utility information
  return action
```

Basic Probability Notation (1)

- What are probabilities about?
  - Assertions about possible worlds: How probable they are
  - Set of all possible worlds: **Sample Space Ω**
    - **Each possible world $\omega \in \Omega$** might have a probability **$P(\omega)$**
    - The possible worlds are mutually exclusive and exhaustive
      - Exhaustiveness can be achieved by adding to a list of possible worlds a further element for the rest called „other“
        - e.g. possible worlds for toothache: Cavity, gumProblem, abscess, other
  - A fully specified probability model associates a numerical probability $P(\omega)$ with each possible world with the following property:

$$\mathbf{0 \le P(\omega) \le 1 \text{ for every } \omega \text{ and } \sum_{\omega \in \Omega} P(\omega) = 1}$$

  - Example: A world with two different dices has 36 possible worlds: (1,1), (1,2), … , (6,6)
  - Usually probabilities relate to sets of possible worlds („events“): **$P(\phi) = \sum_{\omega \in \phi} P(\omega)$**
    - Example: All combinations with sum of dices = 11, i.e.: P(Total=11) = P(5,6) + P(6,5)

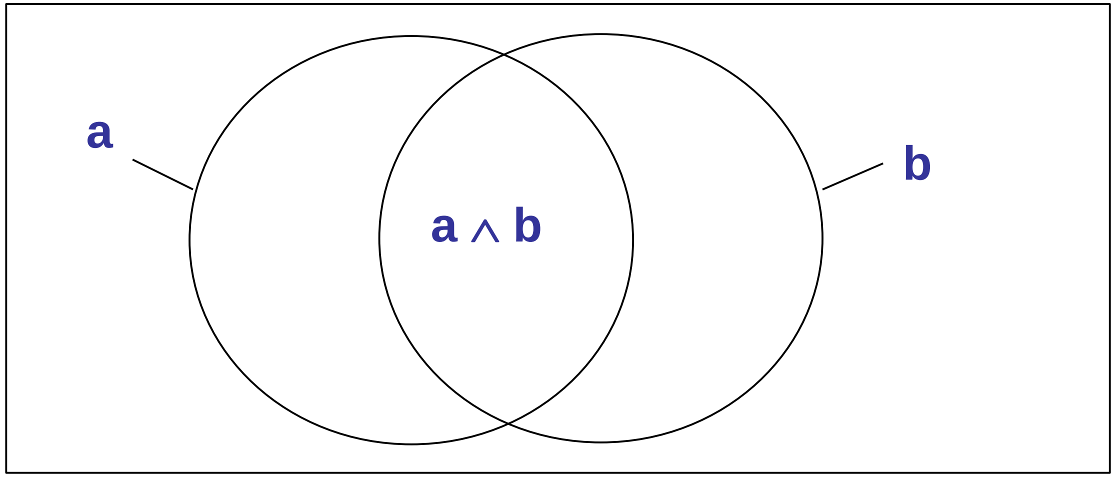
Basic Probability Notation (2)

- In opposition to logical sentences the state of evidence of events can change with new evidence:
  - State before new evidence: **Unconditional (prior) probability**
  - State after new evidence: **Conditional (posterior) probability**
    - Example for unconditional probability: $P(cavity) = 0{,}2$
    - Example for conditional probability: $P(cavity \mid toothache) = 0{,}6$
  - The probability of cavity $P(cavity \mid toothache)$ can change with new information, e.g. if the dentist has excluded cavity: $P(cavity \mid toothache \wedge \neg cavity) = 0{,}001$
- Definition of conditional probabilitities in terms of unconditional probabilities:
  - $\mathbf{P(a \mid b) = P(a \wedge b) / P(b)}$ or
  - $\mathbf{P(a \wedge b) = P(a \mid b)\, P(b)}$ (product rule)
    - If a and b are independant: $\mathbf{P(a \wedge b) = P(a)\, P(b)}$

Probability Axioms

Kolmogorov's axioms:

- $0 \leq P(a) \leq 1$ for every a and $\sum_{a \in \Omega} P(a) = 1$
- $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



Further axioms (can be partially be derived)

- $P(\phi) = \sum_{a \in \phi} P(a)$
- P(True) = 1; P(False) = 0
- $P(a \wedge b) = P(a|b)\, P(b)$ or $P(a|b) = P(a \wedge b) \,/\, P(b)$
- $P(\neg a) = 1 - P(a)$

Probability Distributions

- Variables in probability theory are called random variables
- Random variables have a range
  - **Boolean:** {true, false}, e.g. Toothache
  - **Discrete:** Set of arbitrary tokens, e.g. Weather: {sun, rain, cloud, snow}
  - **Continuous:** Range instead of point value, since a point value has the probability zero. Variable takes on some value x as a parametrized function of x, usually called probability density function, e.g. P(NoonTemperature = x) = Uniform (x; 20C, 25C)
- **Probability distribution** of a random variable: Assignment of probabilities to all possible values, e.g. **P**(Weather) = ⟨0.6, 0.1, 0.29, 0.01⟩ for the ordering ⟨sun, rain, cloud, snow⟩
- **Joint probability distribution** of multiple variables: Probability of all combinations of values of the variables, e.g. **P**(Weather, Toothache): 4 x 2 table of probability values
- **Full joint probability distribution:** Probability of all combinations of random variables of a world, from which all probabilities can be computed

Example for Full Joint (Probability) Distribution

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	0.108	0.012	0.072	0.008
¬ cavity	0.016	0.064	0.144	0.576

… for the world consisting of three boolean variables: toothache, cavity and catch (a test used by dentists): each combination of values for the three variables has a probability.

From the full joint distribution all other probabilities can be computed, e.g.

P(cavity) = 0.108+0.012+0.072+0.008 = 0.2

P(cavity $\wedge$ toothache) = 0.108+0.012 = 0.12

P(cavity|toothache) = P(cavity $\wedge$ toothache) / P(cavity) = 0.12/0.2 = 0.6

Bets According to Probability Theory

- If an agent has some degree of belief in proposition a, the agent should be able to state odds at which it is indifferent to a bet for or against a.
  - Example: Agent A believes „Bayern München will win the next soccer game with probability 0.8“
    - Consistent bet: If BM wins, A gets 1€, if not, A looses 4€
- If another agent has a different degree of belief in the same proposition, and they bet against each other, that agent whose belief is less accurate in reflecting the world will loose money in the long run.

Inconsistent Beliefs

With inconsistent beliefs, an agent would always loose money!

Proposition	Agent 1's belief	Agent 2 bets	Agent 1 bets	Agent 1 payoffs for each outcome			
				a, b	$a, \neg b$	$\neg a, b$	$\neg a, \neg b$
a	0.4	\$4 on a	\$6 on $\neg a$	–\$6	–\$6	\$4	\$4
b	0.3	\$3 on b	\$7 on $\neg b$	–\$7	\$3	–\$7	\$3
$a \vee b$	0.8	\$2 on $\neg(a \vee b)$	\$8 on $a \vee b$	\$2	\$2	\$2	–\$8
				–\$11	–\$1	–\$1	–\$1

- Agent2 offers Agent1 (A1) the following combined bet:
  - If ¬a holds (60%), A1 gets 4€ [utility 0,6*4], otherwise [a; 40%] A1 pays 6€ [utility 0,4*-6]
  - If ¬b holds (70%), A1 gets 3€ [utility 0,7*3], otherwise [b; 30%] A1 pays 7€ [utility 0,3*-7]
  - If a or b holds (80%), A1 gets 2€ [utility 0,8*2], otherwise [¬(a∨b); 20%] A1 pays 8€ [utility 0,2*-8]

➢ Agent1 looses always between 1 and 11 €

Probabilistic Inference

- Computation of posterior probabilities for query propositions given the observed evidence
- From a full joint distribution (see above), all queries can be answered

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	0.108	0.012	0.072	0.008
¬ cavity	0.016	0.064	0.144	0.576

- Computation of apriori-probability of a variable Y:
  - **Marginalization:** Sum of all terms containing Y:
    - General: $P(Y) = \sum_z P(Y, Z=z) = \sum_z P(Y,z)$
    - Example: P(cavity): 0,108 + 0,012 + 0,072 + 0,008 = 0,2
  - **Conditioning:** Using the product rule, we can replace P(Y,z) by P(Y|z)P(z) obtaining:
    - $P(Y) = \sum_z P(Y|z)\, P(z)$
    - Example: P(cavity) =

(P(cavity)|P(catch,toothache)) * P(catch,toothache) +
(P(cavity)|P(¬catch,toothache)) * P(¬catch,toothache) +
(P(cavity)|P(catch,¬toothache)) * P(catch, ¬toothache) +
(P(cavity)|P(¬catch,¬toothache)) * P(¬catch,¬toothache)

- Looks complicated, but is useful for computing conditional probabilities

Computing Conditional Probabilities (1)

- Example:
  - P(cavity|toothache) = P(cavity $\wedge$ toothache) / P(toothache)
    = (0.108 + 0.012) / (0.108 + 0.012 + 0.016 + 0.064) = 0.6
  - P($\neg$cavity|toothache) = P($\neg$cavity $\wedge$ toothache) / P(toothache)
    = (0.016 + 0.064) / (0.108 + 0.012 + 0.016 + 0.064) = 0.4
  - Sum of both values: 1.0 (100%)
- The term P(toothache) is the denominator of both calculations
  - ➢ Normalization constant for the distributation P(Cavity|toothache)
  - ➢ Called **α** : Ensures, that the sum is 1.0
- P(Cavity|toothache) = α **P**(Cavity, toothache)

  = α [**P**(Cavity, toothache, catch) + **P**(Cavity, toothache, $\neg$catch)]

  = α [$\langle$0.108,0.016$\rangle$+ $\langle$0.012,0.064$\rangle$] = α $\langle$0.12,0.08$\rangle$ = $\langle$0.6,0.4$\rangle$

	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

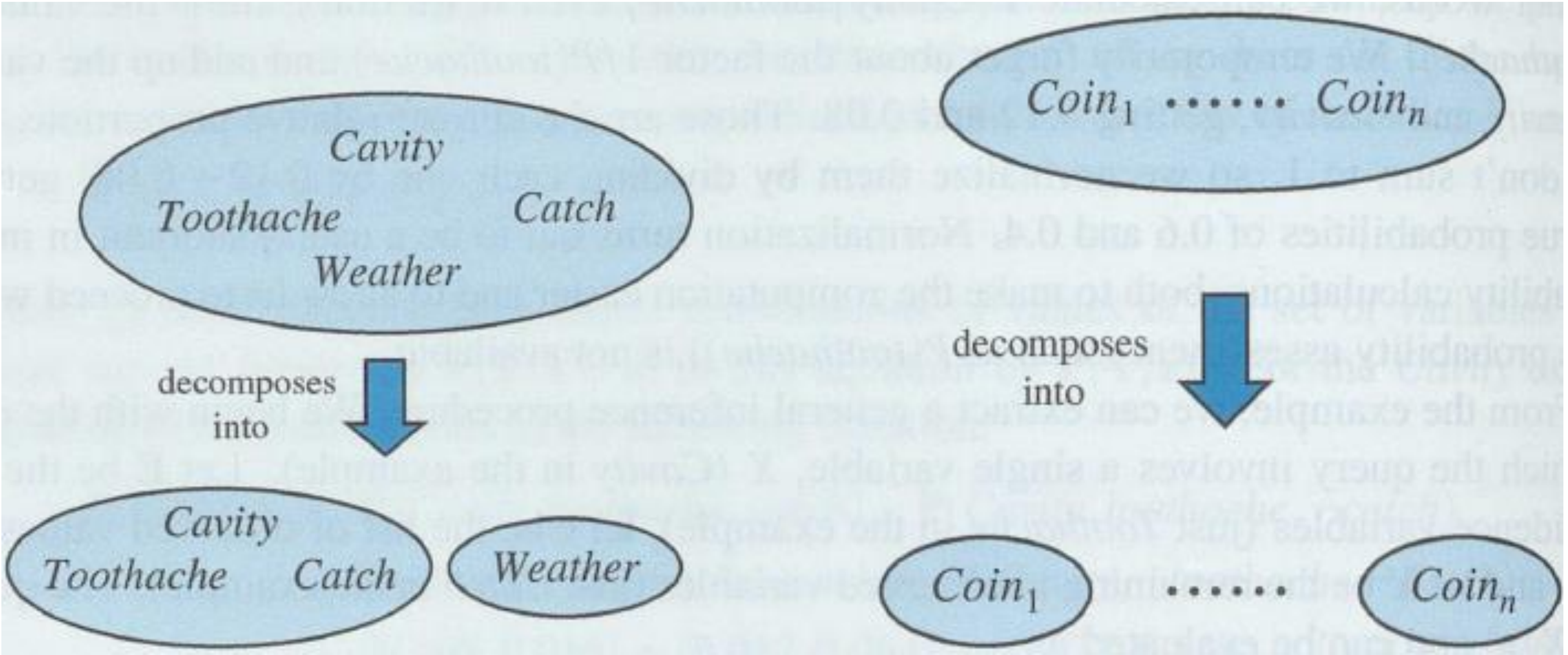
Computing Conditional Probabilities (2)

- General rule: $\mathbf{P}(X|\mathbf{e}) = \alpha\, \mathbf{P}(X,\mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X,\mathbf{e},\mathbf{y})$ where
  - **X** be a single variable (e.g. cavity),
  - **E** the list of evidence variables and **e** the list of observed values for them (e.g. toothache)
  - **Y** the list of the remaining unobserved variables with values **y** (e.g. catch)
- X, **E, Y**: complete set of variables in the domain
- P(X,**e,y**): subset of probabilities from the full joint distribution

- Drawback: Exponential time (2^n) with n boolean variables
  - Similar to inference with truth tables for propositional logic
  - We need more efficient inference procedures (like DPLL or WalkSAT)

Independance

- If two variables are independant, many equations can be simplified:
  - P(a|b) = P(a)
  - P(a $\wedge$ b) = P(a) P(b)
- Examples:
  - Weather is independant from dental problems
  - Coin flips are independant



- Concrete examples:
  - P(cloud|toothache, catch, cavity) = P(cloud)
  - P(toothache, catch, cavity, cloud) = P(cloud) P(toothache, catch, cavity)
  - General: **P**(Toothache, Catch, Cavity, Weather) = **P**(Toothache, Catch, Cavity) **P(**Weather)

Bayes' Rule

- $P(b|a) = P(a|b)\,P(b) / P(a)$
  - Justification:
    - $P(a \wedge b) = P(a|b)\,P(b)$
    - $P(b \wedge a) = P(b|a)\,P(a)$

$$\left.\begin{array}{l} P(a \wedge b) = P(a|b)\,P(b) \\ P(b \wedge a) = P(b|a)\,P(a) \end{array}\right\} P(b|a)\,P(a) = P(a|b)\,P(b) \Rightarrow P(b|a) = P(a|b)\,P(b) / P(a)$$

- Application:
  - $P(cause|effect) = P(effect|cause)\,P(cause) / P(effect)$
  - $P(disease|symptom) = P(symptom|disease)\,P(disease) / P(symptom)$
    - $P(meningitis|stiff\_neck) = P(stiff\_neck|meningitis)\,P(meningitis) / P(stiff\_neck)$
    - Assume: $P(meningitis) = 10^{-5}$; $P(stiff\_neck) = 0.05$; $P(stiff\_neck|meningitis) = 0.5$
      - $P(meningitis|stiff\_neck) = 10^{-5} * 0.5 / 0.05 = 10^{-4}$ (factor 10 more probable)
- Why not use probabilities P(disease|symptom) directly?
  - They are less stable than P(symptom|disease), e.g. may change with seasons.

Bayes‘ Rule: Combining Evidence

- Example: $\mathbf{P}(Cavity|toothache \wedge catch)$
  - Computation with full joint distribution: $\alpha\langle 0.108, 0.016\rangle \approx \langle 0.871, 0.129\rangle$
  - Computation with Bayes‘ rule: $\alpha\ \mathbf{P}(toothache \wedge catch|Cavity)\ \mathbf{P}(Cavity)$
    - What is $\mathbf{P}(toothache \wedge catch|Cavity)$? – Probably not easily available!
  - However, if catch and cavity would be independant, we can use the product rule
    - They are not independant since both are caused by cavity, but conditional independance is sufficient:
    - $\mathbf{P}(toothache \wedge catch|Cavity) = \mathbf{P}(toothache|Cavity)\ \mathbf{P}(catch|Cavity)$
- General definition of conditional independance: $\mathbf{P}(X,Y|Z) = \mathbf{P}(X|Z)\ \mathbf{P}(Y|Z)$
- The full joint distributation: $\mathbf{P}(Cause|Effect_1, \ldots, Effect_n) = P(Cause)\ \Pi_i\ \mathbf{P}(Effect_i|Cause)$
- Also called „naive Bayes“, because the variables are often not strictly independant
- Bayes‘ Rule: $\mathbf{P}(Cause|\mathbf{e}) = \alpha\ \mathbf{P}(Cause)(\Pi_j\ \mathbf{P}(e_j|Cause)$

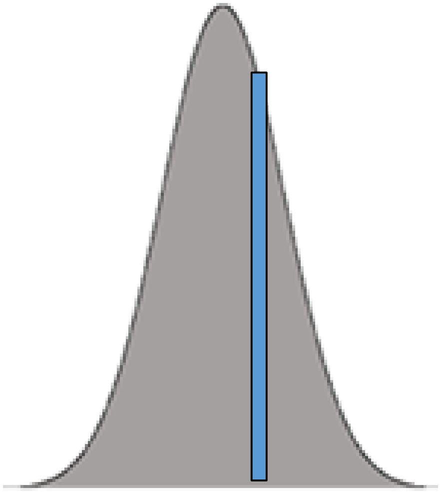
Bayes Rule

- Bayes' Rule (from last slide): **P**(Cause|**e**) = α **P**(Cause) Π_j **P**(e_j|Cause)
- Reformulated for inferring a diagnosis D from S_1 … S_n symptoms:
  - **P**(D|S_1, … , S_n) = α P(D) * P(S_1|D) * … * P(S_n|D)
- If we want to decide the probability of a diagnosis D_i from a set an exclusive set of diagnoses, α is the sum of the probabilities of all diagnoses (normalization constant):
  - P(D_i|S_1, … , S_m) = α P(D_i) * P(S_1|D_i) * … * P(S_m|D_i)
  - or

$$P(D_i / S_1 \& \ldots \& S_m) = \frac{P(D_i) * P(S_1 / D_i) * \ldots * P(S_m / D_i)}{\sum_{j=1}^{n} P(D_j) * P(S_1 / D_j) * \ldots * P(S_m / D_j)}$$

Numerical (Continuous) Variables

- Discrete and numerical variables can easily combined with Bayes‘ rule
  - e.g. temperature = 25 Celsius and weather = sunny
- Solution1: Discretization of numerical variables
  - Needs knowledge about threshold; loss in precision
  - e.g. <0 = icy, [0,10] = cold,]10,20] = medium;]20,30] = warm; >30 = hot
- Solution2: Direct use by a probability density function
  - Needs knowledge about density function
  - Standard density function: Normal distribution with mean (μ) and standard derivation (σ)
    - Compute μ and σ for each numerical attribute (symptom) and for each class (e.g. diagnosis) separately
  - For a new case map its value x (e.g. temp. = 25 C) to a small area: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
  - The result can be treated like a probability in Bayes‘ rule



Origin of Probabilities

- From experiments and statistical computations
- From knowlege
- From expectations

Examples: What is the probability

… that a playing card in Skat is jack of clubs?

… that the sun rises tomorrow?

… that meningitis causes a stiff neck?

Computation of Probabilities in Data Sets

Table 1.2 The weather data.

Outlook	Temperature	Humidity	Windy	Play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

Example from:
Witten, I., Frank, E., Hall, M., Pal, C.: Data Mining, 4. edition, 2017.

The conditional probabiliy for „Outlook=sunny“ under the assumption „Play = yes“ is P(sunny|Play=yes) = 2/9

Table 4.2 The weather data with counts and probabilities.

Outlook			Temperature			Humidity			Windy			Play	
	yes	no		yes	no		yes	no		yes	no	yes	no
sunny	2	3	hot	2	2	high	3	4	false	6	2	9	5
overcast	4	0	mild	4	2	normal	6	1	true	3	3		
rainy	3	2	cool	3	1								
sunny	2/9	3/5	hot	2/9	2/5	high	3/9	4/5	false	6/9	2/5	9/14	5/14
overcast	4/9	0/5	mild	4/9	2/5	normal	6/9	1/5	true	3/9	3/5		
rainy	3/9	2/5	cool	3/9	1/5								

The Apriori-probability for „Play = yes“ is: P(Play=yes) = 9/14.

Forecast Example: A New Day

- A new day:

Outlook = sunny

Temperature = cool

Humidity = high

Windy = true

Play = ???

Table 4.2 The weather data with counts and probabilities.

Outlook			Temperature			Humidity			Windy			Play	
	yes	*no*		*yes*	*no*		*yes*	*no*		*yes*	*no*	*yes*	*no*
sunny	2	3	hot	2	2	high	3	4	false	6	2	9	5
overcast	4	0	mild	4	2	normal	6	1	true	3	3		
rainy	3	2	cool	3	1								
sunny	2/9	3/5	hot	2/9	2/5	high	3/9	4/5	false	6/9	2/5	9/14	5/14
overcast	4/9	0/5	mild	4/9	2/5	normal	6/9	1/5	true	3/9	3/5		
rainy	3/9	2/5	cool	3/9	1/5								

P (play=no|sunny, cool, high, true) = P (no) * P(sunny/no) * P(cool/no) * P(high/no) * P(true/no)
= 5/14 * 3/5 * 1/5 * 4/5 * 3/5 = 0.0206

P (play=yes|sunny, cool, high, true) = P(yes) * P(sunny/yes) * P(cool/yes) * P(high/yes) * P(true/yes)
= 9/14 * 2/9 * 3/9 * 3/9 * 3/9 = 0.0053

P(Play|sunny, cool, high, true) **= α ⟨0.0206, 0.0053⟩ = ⟨0.795, 0.205⟩**

α = 1/(0,0206 + 0,0053) = 1/0,0259 ⇒ P (play=no | sunny, cool, high, true) = 0,0206/0,0259 = 79,5%

P (play=yes| sunny, cool, high, true) = 0,0053/0,0259 = 20,5%

Forecast with Numerical Probabilities in Data Sets

Outlook			Temperature			Humidity			Windy			Play	
	Yes	No		Yes	No		Yes	No		Yes	No	Yes	No
Sunny	2	3		83	85		86	85	False	6	2	9	5
Overcast	4	0		70	80		96	90	True	3	3		
Rainy	3	2		68	65		80	70					
				64	72		65	95					
				69	71		70	91					
				75			80						
				75			70						
				72			90						
				81			75						
Sunny	2/9	3/5	*Mean*	73	74.6	*Mean*	79.1	86.2	False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	*Std dev*	6.2	7.9	*Std dev*	10.2	9.7	True	3/9	3/5		
Rainy	3/9	2/5											

Another new day:

Outlook = **sunny**
Temperature = **66**
Humidity = **90**
Windy = **true**
Play = **???**

f(Temperature=66/yes) = 0.0340 = $\frac{1}{\sqrt{2\pi}\cdot 6.2} e^{-\frac{(66-73)^2}{2\cdot 6.2^2}}$ $\left(f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)$

f(Humidity = 90/yes) = 0.0221

f(Temperature=66/no) = 0.0279

f(Humidity = 90/no) = 0.0381

Likelihood of **Play=yes**: 2/9 * 0.0340 * 0.0221 * 3/9 * 9/14 = 0.000036 (**20,8%**)

Likelihood of **Play = no**: 3/5 * 0.0279 * 0.0381 * 3/5 * 5/14 = 0.000137 (**79,2%**)

Text Classification with Naive Bayes

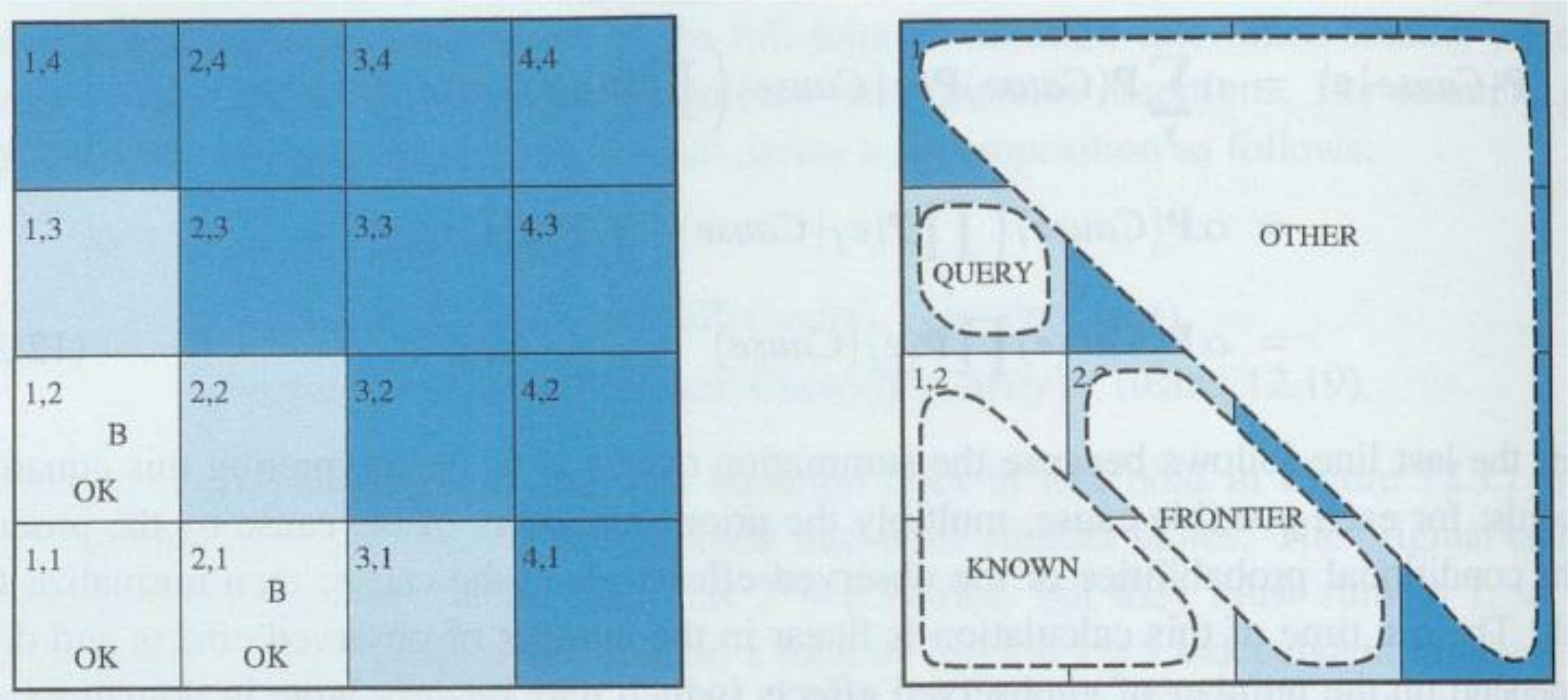
- Goal: Classify a text unit (sentence, paragraph, news item) as belonging to a category, e.g. to a major section of a newspaper like news, sports, business, weather, entertainment
- Method: Compute the probability of each word for each category from a training corpus
- Two example sentences:
  - *Stocks rallied on Monday, with major indexes gaining 1% as optimism persisted over the first quarter earnings season.*
  - *Heavy rain continued to pound much of the east cost on Monday, with flood warnings issued in New York City and other locations.*
- Compute apriori probabilities: P(weather), P(business) etc.
- Compute conditional probabilities:
  - P(rain|weather), P(rain|business), P(stock|weather), P(stock|business), etc.
- Improvements: Ignore „stop words“ (very frequent words), normalize words by stemming (reduce a word to its stem avoiding singular/plural variations etc.), ignore very rare words, etc.

Wumpus World Revisited

Computate probability for unsafe squares to choose the least probable, if no safe squares exist

Example:

- Situation after finding a breeze in both [1,2] and [2,1] (left)
- Division of squares in Known, Frontier and Other for a Query about [1,3] (right)

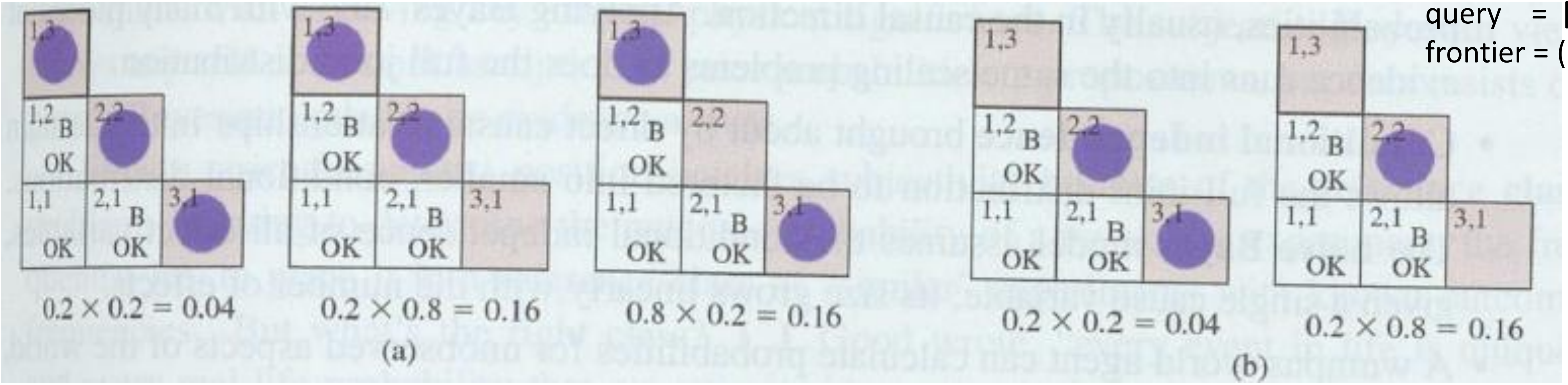


Approach 1: Computation based on the full joint probability: Possible, but expensive

Approach 2: Computation based on logical models of possible distributation of pits and independance assumption from Other to Frontier and Query.

Logical Models of Distribution of Pits

query = [1,3];
frontier = ([2,2], [3,1])



(a) Three possible distributions of pits in frontier, if [1,3] contains a pit, weighted with probability 0.2

(b) Two possible distributions of pits in frontier, if [1,3] does not contain a pit, weighted with probability 0.8

All probabilities are 0.2 for pit and 0.8 for no pit in each square:

(a) 0.2 * (0.2*0.2 + 0.2*0.8 + 0.8*0.2) = 0.2*0.36 = 0.072

(b) 0.8 * (0.2*0.2 + 0.2*0.8) = 0.8*0.2 = 0.16

Normalization: $\alpha \langle 0.072, 0.16 \rangle \approx \langle 0.31, 0.69 \rangle \Rightarrow$ **The probability of a pit in [1,3] is ca. 31% (and also in [3,1])**

*Four possible distributions for pit in [2,2]: 0.2(0,8*0.8+0.8*0.2+0.2*0.8+0.2*0.2)=0.2; One distribution for no pit in [2,2]=0.8(0.2*0.2)=0.032; Normalization:* $\alpha \langle 0.2, 0.032 \rangle \approx \langle 0.86, 0.14 \rangle \Rightarrow$ ***The probability of a pit in [2,2] is ca. 86%***

Discussion

- The main assumption of naive Bayes, the conditional independance of the symptoms (effects) given the diagnosis (cause) is usually not correct. Nevertheless, it produces often good results and is very efficient. Therefore, it is often used as **baseline**.
- **Hints for practical use:**
  - Ignore attributes being redundant to other attributes or aggregate them (e.g. *height* in meter and *weight* in kg to *body mass index* by (*weight/height*^{2})
  - Ignore attributes being obviously irrelevant (e.g. identifier)
  - To avoid 0-probabilities, initialize the count of all attribute-values with 1
  - Numerical (continuous) attributes can be used directly without need of discretization

- **Improvement by boosting and voting**:
  - Use several hypotheses with voting
  - Generation of different hypotheses e.g. by weighting of falsely classified cases (equivalent to multiplication of cases)