

From  
Standard Software for  
Enterprise Resource Planning

To  
Foundations of Machine Learning  
and Data Mining

Lecturer: Rainer Marrone, Prof. Dr. Ralf Möller  
Lab classes: Rainer Marrone

Lecture: Thursdays (90 minutes)  
Lab classes: Fridays (60 minutes)

Prerequisite:  
Basic Mathematics

# Recap

Probability is a rigorous formalism for uncertain knowledge

Joint probability distribution specifies probability of every atomic event

Queries can be answered by summing over atomic events

For nontrivial domains, we must find a way to reduce the joint size

Independence and conditional independence provide the tools

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

$$\begin{aligned}
 P(\text{cavity}|\text{toothache}) &= P(\text{cavity} \wedge \text{toothache}) / P(\text{toothache}) \\
 &= (0.108 + 0.012) / (0.108 + 0.012 + 0.016 + 0.064) = 0.6
 \end{aligned}$$

Interpretation: After observing toothache, the patient is no longer an "average" one, and the prior probabilities of cavity is no longer valid

$P(\text{Cavity}|\text{Toothache})$  is calculated by keeping the ratios of the probabilities of the 4 cases unchanged, and normalizing their sum to 1

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

$$P(\text{Cavity} | \text{Toothache}) = P(\text{Cavity} \wedge \text{Toothache}) / P(\text{Toothache})$$

$$= (0.108 + 0.012) / (0.108 + 0.012 + 0.016 + 0.064) = 0.6$$

$$P(\neg \text{Cavity} | \text{Toothache}) = P(\neg \text{Cavity} \wedge \text{Toothache}) / P(\text{Toothache})$$

$$= (0.016 + 0.064) / (0.108 + 0.012 + 0.016 + 0.064) = 0.4$$

$$P(c | \text{Toothache}) = \alpha P(c \wedge \text{Toothache})$$

$$= \alpha \sum_{pc} P(c \wedge \text{Toothache} \wedge pc)$$

$$= \alpha [(0.108, 0.016) + (0.012, 0.064)]$$

$$= \alpha (0.12, 0.08) = (0.6, 0.4)$$

normalization constant

# Conditional Probability

- $P(A \wedge B) = P(A|B) P(B)$   
 $= P(B|A) P(A)$
- $P(A \wedge B \wedge C) = P(A|B, C) P(B \wedge C)$   
 $= P(A|B, C) P(B|C) P(C)$
- $P(\text{Cavity}) = \sum_t \sum_{pc} P(\text{Cavity} \wedge t \wedge pc)$   
 $= \sum_t \sum_{pc} P(\text{Cavity}|t, pc) P(t \wedge pc)$
- $P(c) = \sum_t \sum_{pc} P(c \wedge t \wedge pc)$   
 $= \sum_t \sum_{pc} P(c|t, pc) P(t \wedge pc)$

# Independence

- Two random variables  $A$  and  $B$  are **independent** if

$$P(A \wedge B) = P(A) P(B)$$

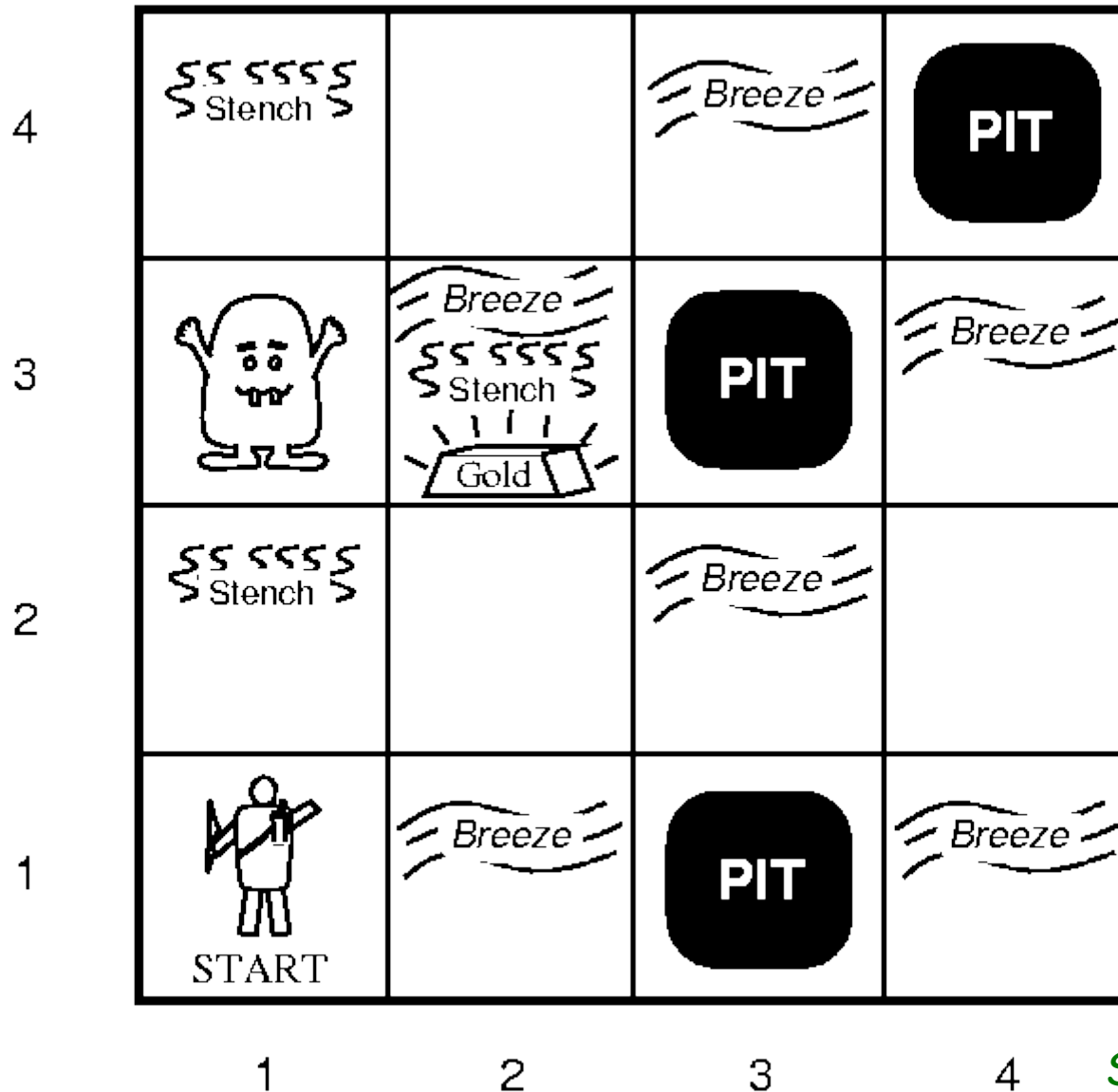
hence if  $P(A|B) = P(A)$

- Two random variables  $A$  and  $B$  are **independent given  $C$** , if

$$P(A \wedge B | C) = P(A | C) P(B | C)$$

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

# Wumpus World



Wumpus: Monster  
(death)

Pit: (death)

Stench: (near Wumpus)

Breeze: (near pit)

Glitter: (goal)

Hit: Agent moves  
against wall

Agent can move, turn,  
and shoot one arrow in  
the the direct it heads to

Agent has one arrow  
that kills the Wumpus  
(scream)

# Wumpus World (2)

[1, 2] und [2, 1] sind sicher:

1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2 OK	2,2	3,2	4,2
1,1 A OK	2,1 OK	3,1	4,1

(a)

- A** = Agent
- B = Breeze
- G = Glitter, Gold
- OK = Safe square
- P = Pit
- S = Stench
- V = Visited
- W = Wumpus

1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2 OK	2,2 P?	3,2	4,2
1,1 V OK	2,1 A B OK	3,1 P?	4,1

(b)

Slide: St. Russell

# Wumpus Welt (3)

1,4	2,4	3,4	4,4
1,3 W!	2,3	3,3	4,3
1,2 <b>A</b> S OK	2,2 OK	3,2	4,2
1,1 V OK	2,1 B V OK	3,1 P!	4,1

(a)

**A** = Agent  
 B = Breeze  
 G = Glitter, Gold  
 OK = Safe square  
 P = Pit  
 S = Stench  
 V = Visited  
 W = Wumpus

1,4	2,4 P?	3,4	4,4
1,3 W!	2,3 <b>A</b> S G B	3,3 P?	4,3
1,2 S V OK	2,2 V OK	3,2	4,2
1,1 V OK	2,1 B V OK	3,1 P!	4,1

(b)

# Probabilistic Wumpus World

1,4	2,4	3,4	4,4
1,3	2,3	3,3	4,3
1,2 <b>B</b> <b>OK</b>	2,2	3,2	4,2
1,1 <b>OK</b>	2,1 <b>B</b> <b>OK</b>	3,1	4,1

$P_{ij} = true$  iff  $[i, j]$  contains a pit

$B_{ij} = true$  iff  $[i, j]$  is breezy

Include only  $B_{1,1}$ ,  $B_{1,2}$ ,  $B_{2,1}$  in the probability model

Slide: St. Russell

# Specifying the probability model

The full joint distribution is  $\mathbf{P}(P_{1,1}, \dots, P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$

Apply product rule:  $\mathbf{P}(B_{1,1}, B_{1,2}, B_{2,1} \mid P_{1,1}, \dots, P_{4,4})\mathbf{P}(P_{1,1}, \dots, P_{4,4})$

(Do it this way to get  $P(\textit{Effect} \mid \textit{Cause})$ .)

First term: 1 if pits are adjacent to breezes, 0 otherwise

Second term: pits are placed randomly, probability 0.2 per square:

$$\mathbf{P}(P_{1,1}, \dots, P_{4,4}) = \prod_{i,j=1,1}^{4,4} \mathbf{P}(P_{i,j}) = 0.2^3 \times 0.8^{16-3}$$

for 3 pits.

# Observations and query

We know the following facts:

$$b = \neg b_{1,1} \wedge b_{1,2} \wedge b_{2,1}$$

$$known = \neg p_{1,1} \wedge \neg p_{1,2} \wedge \neg p_{2,1}$$

Query is  $\mathbf{P}(P_{1,3}|known, b)$

Define *Unknown* =  $P_{ij}$ s other than  $P_{1,3}$  and *Known*

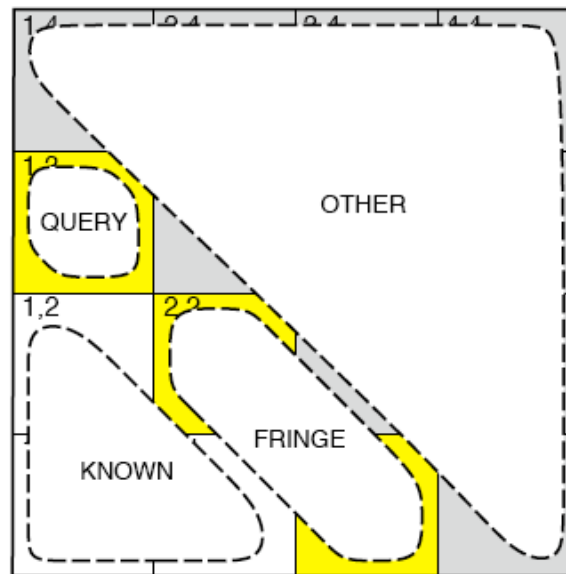
For inference by enumeration, we have

$$\mathbf{P}(P_{1,3}|known, b) = \alpha \sum_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b)$$

Grows exponentially with number of squares!

# Using conditional independence

Basic insight: observations are conditionally independent of other hidden squares given neighbouring hidden squares



Define  $Unknown = Fringe \cup Other$

$$\mathbf{P}(b|P_{1,3}, Known, Unknown) = \mathbf{P}(b|P_{1,3}, Known, Fringe)$$

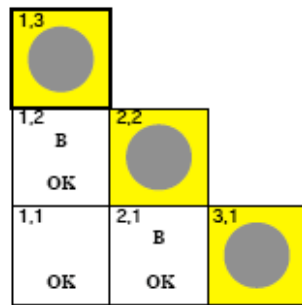
Manipulate query into a form where we can use this!

Slide: St. Russell

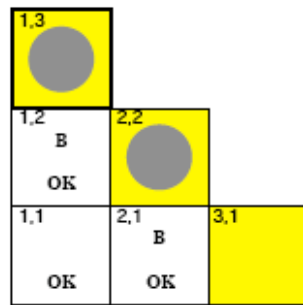
# Using conditional independence (2)

$$\begin{aligned}\mathbf{P}(P_{1,3}|known, b) &= \alpha \sum_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b) \\ &= \alpha \sum_{unknown} \mathbf{P}(b|P_{1,3}, known, unknown) \mathbf{P}(P_{1,3}, known, unknown) \\ &= \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe, other) \mathbf{P}(P_{1,3}, known, fringe, other) \\ &= \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe) \mathbf{P}(P_{1,3}, known, fringe, other) \\ &= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3}, known, fringe, other) \\ &= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3}) P(known) P(fringe) P(other) \\ &= \alpha P(known) \mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) P(fringe) \sum_{other} P(other) \\ &= \alpha' \mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) P(fringe)\end{aligned}$$

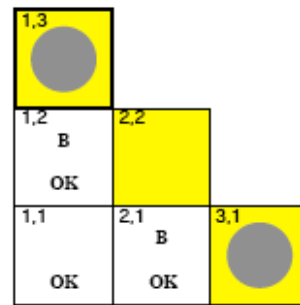
# Using conditional independence (3)



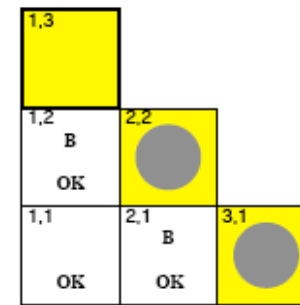
$$0.2 \times 0.2 = 0.04$$



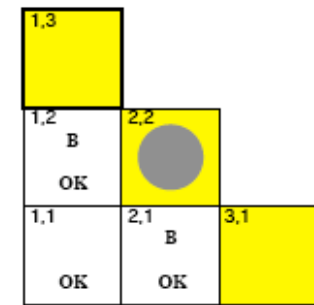
$$0.2 \times 0.8 = 0.16$$



$$0.8 \times 0.2 = 0.16$$



$$0.2 \times 0.2 = 0.04$$



$$0.2 \times 0.8 = 0.16$$

$$\mathbf{P}(P_{1,3} | \textit{known}, b) = \alpha' \langle 0.2(0.04 + 0.16 + 0.16), 0.8(0.04 + 0.16) \rangle \\ \approx \langle 0.31, 0.69 \rangle$$

$$\mathbf{P}(P_{2,2} | \textit{known}, b) \approx \langle 0.86, 0.14 \rangle$$