Università di Bergamo
Facoltà di Ingegneria

Intelligenza Artificiale

Paolo Salvaneschi

A9_1 V1.9

Trattamento dell’Incertezza

Nota: è utilizzato in parte il materiale didattico associato al testo di Stuart J. Russell, Peter Norvig

INDICE

- Incertezza
- Probabilità
- Assiomi della probabilità
- Inferenza probabilistica per enumerazione
- Indipendenza
- Regola di Bayes
- Utilizzo della regola di Bayes

Incertezza

- Azione per un agente logico: condurre un passeggero all’aeroporto in tempo per prendere un volo
- Piano: partenza 90 minuti prima del volo, velocità di guida normale
- Problemi:
  - Osservabilità parziale (stato della strada, piani degli altri guidatori,…)
  - Eventi inaspettati (si buca una gomma,…)
  - Sensori affetti da rumore (stato del traffico via radio)
  - Immensa complessità del modello di predizione del traffico

- Cfr: Situation Calculus - the Qualification Problem

Incertezza

- Piano: partenza 90 minuti prima del volo, velocità di guida normale
- Il piano mi porterà in tempo se non ci sono incidenti sul percorso, se non buco una gomma, se…..
- Un agente puramente logico non è in grado di generare un piano che garantisca il successo.

- Trattare la conoscenza incerta
- Trattare piani diversi che hanno vantaggi e svantaggi (es. partire prima e aspettare a lungo in aeroporto)

- Conoscenza incerta e logica

Verso diagnostico

$$\forall p \; Symptom\ (p, Toothache) \Rightarrow Disease\ (p, Cavity)$$

La regola è scorretta. I mal di denti non implica necessariamente la carie

$$\begin{aligned}\forall p \; Symptom\ (p, Toothache) \Rightarrow\ & Disease\ (p, Cavity) \quad \vee \\ & Disease\ (p, Abscess) \; \vee \\ & \dots\dots\dots\dots\dots\dots\end{aligned}$$

Quanto è grande la lista delle cause?

Verso causale

$$\forall p \; Disease\ (p, Cavity) \Rightarrow Symptom\ (p, Toothache)$$

Cambiamo direzione. Passiamo da una regola diagnostica ad una regola causale. Anche questa non è corretta.
Es. ci sono carie che non causano mal di denti

Incertezza

- Conoscenza incerta e logica
- Usare FOL per questo problema di diagnosi fallisce:
  - Costo di costruzione di una lista completa di antecedenti e conseguenti
  - Conoscenza teorica incompleta
  - Conoscenza pratica incompleta (non si hanno tutte le possibili misure)
- Un agente dispone al meglio di un grado di credenza a proposito di una affermazione

Incertezza

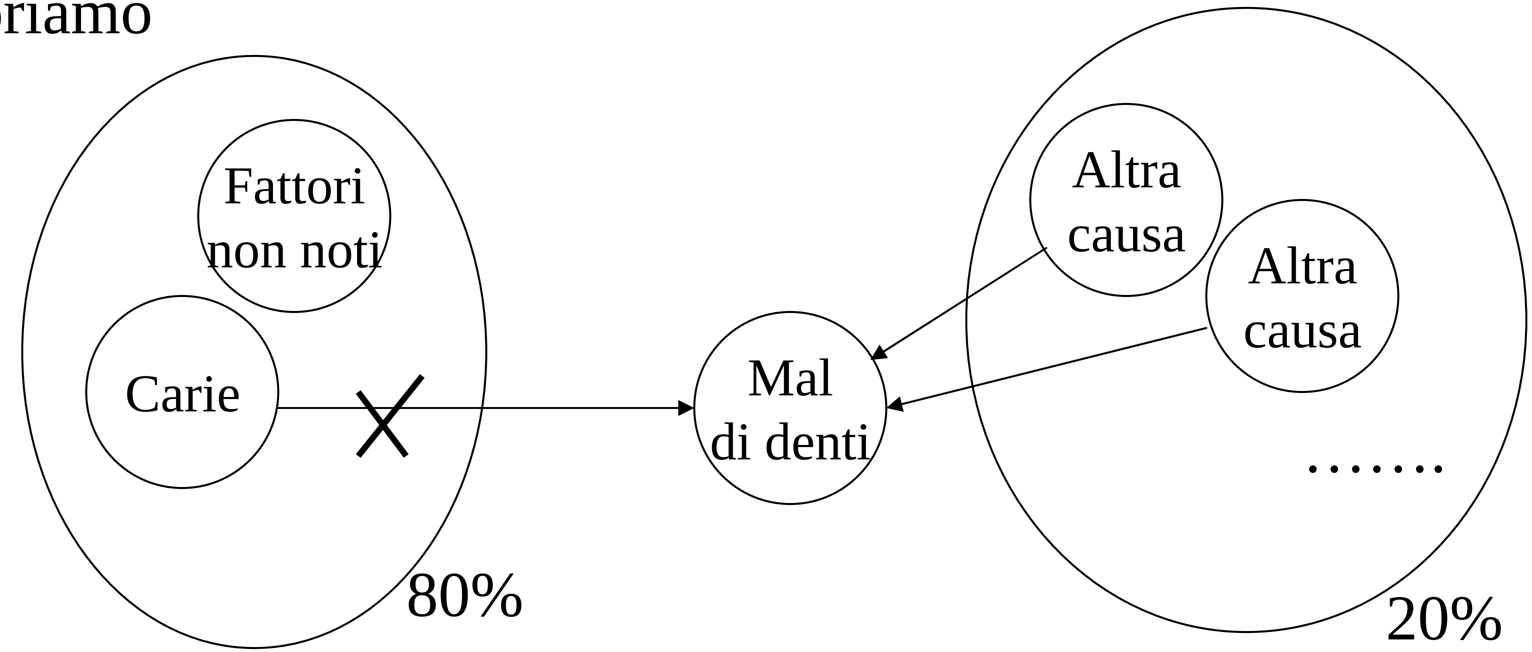
- Modo per riassumere gli effetti di mancanza di conoscenza: Probabilità

  Con una probabilità dell'80% il paziente ha una carie se ha mal di denti
- L'80% riassume tutti i fattori richiesti affinchè una carie causi il mal di denti e il caso in cui ci sia mal di denti e carie e i due fenomeni siano sconnessi.
- Il 20% riassume tutte le altre possibili cause di mal di denti che ignoriamo

- Conoscenza derivata ad es. da dati statistici

Incertezza

- L’80% riassume tutti i fattori richiesti affinchè una carie causi il mal di denti e il caso in cui ci sia mal di denti e carie e i due fenomeni siano sconnessi.
- Il 20% riassume tutte le altre possibili cause di mal di denti che ignoriamo



Nota: differenza con FOL; closed world assumption

Incertezza

- La probabilità rappresenta una grado di credenza che l’agente ha rispetto a un fatto
  - Il fatto in sè resta vero o falso. L’ontologia non cambia.
  - Probabilità: gradi di credenza

    Fuzzy logic: gradi di verità dei fatti

Incertezza

- Il grado di credenza dipende dalle evidenze (percezioni) raccolte dall'agente
  - Il grado di credenza su un fatto può cambiare se nuovi fatti (evidenze) sono aggiunti alla base di conoscenza (N.B. il valore di verità del fatto non cambia)
  - Probabilità a priori (non condizionata). Evidenza prima di avere percezioni
  - Probabilità a posteriori (condizionata). Evidenza dopo percezioni

Nota: monotonicità ?

Incertezza

- "Trattare piani diversi che hanno vantaggi e svantaggi (es. partire prima e aspettare a lungo in areoporto)"
- Possibili credenze:

Date certe condizioni

$P(A_{25}$ gets me there on time | …) = 0.04

$P(A_{90}$ gets me there on time | …) = 0.70

$P(A_{120}$ gets me there on time | …) = 0.95

$P(A_{1440}$ gets me there on time | …) = 0.9999

- Quale azione scegliere?

  Dipende dalle preferenze dell'agente (es. rischiare di perdere l'aereo vs aspettare a lungo)

  – <u>Utility theory</u> utilizzata per rappresentare e ragionare sulle preferenze

  – <u>Decision theory</u> = <u>probability theory</u> + <u>utility theory</u>

Probabilità

- Linguaggio formale per rappresentare e ragionare con la conoscenza incerta:

- Applicazione della probabilità alla logica delle proposizioni
- Distinzione tra probabilità a priori e condizionata (a posteriori)

- Sintassi:……………

- Proposizioni

- Elemento di base: variabile random
- Come simbolo proposizionale in PL
  - Es. *Cavity* (maiuscolo) (*)
  - Semantica come in PL: mondi possibili definiti assegnando valori a variabili random

(*) *Cavity* variabile random (maiuscolo)

a variabile random non nota (minuscolo) es $P(a) = 1 - P(\neg a)$

cavity abbreviazione di $Cavity = true$

- Ogni variabile random ha un dominio
  - Booleano

    Cavity *<true, false>* *Cavity = true* *Cavity=false*

    cavity $\neg$ *cavity*
  - Discreto

    Weather <sunny,rainy,cloudy,snow> *Weather = rainy*

    valori scritti in minuscolo

    I valori del dominio devono essere esaustivi e mutamente esclusivi
  - Continuo valori = numeri reali (intervalli)

    $X = 4{,}16 \quad X \leq 17{,}6$

- Le proposizioni elementari sono costruite mediante assegnamenti di valori a variabili random

$$Cavity = false$$

$$Weather = rainy$$

- Le proposizioni complesse sono formate da proposizioni elementari e connettivi logici standard

$$Cavity = true \wedge \; Toothache = false$$

$$cavity \wedge \neg\, toothache$$

- Eventi atomici (o campioni)
- Evento atomico: una specificazione completa dello stato del mondo relativamente al quale l'agente ha conoscenza incerta
- Es. Il mondo consiste solo di due variabili booleane *Cavity* e *Toothache*

  Ci sono 4 distinti eventi atomici:

  $$Cavity = false \wedge Toothache = false$$
  $$Cavity = false \wedge Toothache = true$$
  $$Cavity = true \wedge Toothache = false$$
  $$Cavity = true \wedge Toothache = true$$

- Gli eventi Atomici sono esaustivi e mutamente esclusivi

Probabilità

- Probabilità a priori
- $P(a)$ Probabilità a priori di una proposizione a

$P(Cavity = true) = 0.1 \qquad P(cavity) = 0.1$

usata prima dell'arrivo di ogni (nuova) evidenza

(grado di credenza in assenza di ogni altra informazione)

- $\mathbf{P}(a)$ Le probabilità di tutti i possibili valori di una variabile random a

$\mathbf{P}(Weather) = <0.72, 0.1, 0.08, 0.1>$

Vettore di valori.

Sostituisce le equazioni:

$P(Weather=sunny) = 0.72$
$P(Weather=rainy) = 0.1$
……

Distribuzione di probabilità a priori della variabile casuale *Weather* (normalizzata; somma 1)

- ***P*** *(a, b)* Joint probability distribution per l'insieme delle variabili random *a* e *b*
  - Denota la probabilità di tutte le combinazioni di valori delle variabili random

    P*(Weather,Cavity)* matrice di 4 × 2 valori

$Weather =$	$sunny$	$rainy$	$cloudy$	$snow$
$Cavity = true$	0.144	0.02	0.016	0.02
$Cavity = false$	0.576	0.08	0.064	0.08

- Full Joint probability distribution
  - Joint probability distribution per tutte le variabili che sono usate per descrivere il mondo
  - Specifica la probabilità di ogni evento atomico
  - Completa specificazione dell'incertezza sul mondo in questione.
  - Ogni query probabilistica può essere risposta a partire da essa.(vedi più avanti "inferenza per enumerazione")

Probabilità

- Tutto ciò per variabili discrete

- Per variabili continue: Funzioni di densità di probabilità

Probabilità

- Probabilità condizionate (a posteriori)

- $P\ (a \mid b)$ La probabilità di a, stante che tutto ciò che conosco è b

  $P(cavity \mid toothache) = 0.8$

  l'unico sintomo noto del paziente è il mal di denti

- La probabilità a priori può essere definita come un caso particolare di probabilità condizionata (senza evidenze)

  $P(cavity \mid\ )$

- La probabilità condizionata può essere definita nei termini di probabilità incondizionate

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)} \quad if\ P(b) > 0$$

o equivalentemente

$P(a \wedge b) = P(a \mid b)\, P(b)$ (product rule)

$P(a, b)$

Oppure

$P(a \wedge b) = P(b \mid a)\, P(a)$

- Probabilità condizionate
- Notazione ***P*** $\mathbf{P}(X \mid Y)$ (un vettore di valori)
  Fornisce i valori di $P(X = x_i \mid Y = y_j)$ per ogni i e j

$$\mathbf{P}(Weather, Cavity) = \mathbf{P}(Weather \mid Cavity)\, \mathbf{P}(Cavity)$$

Notazione compatta che sostituisce 4 x 2 equazioni:

$P(Weather=sunny, Cavity=true) = P(Weather=sunny \mid Cavity=true)\ P(Cavity=true)$
$P(Weather= rainy, Cavity= true) = P(Weather= rainy \mid Cavity= true)\ P(Cavity=true)$
$P(Weather=cloudy, Cavity=true) = P(Weather=cloudy \mid Cavity=true)\ P(Cavity=true)$
$P(Weather=snow, Cavity=true) = P(Weather=snow \mid Cavity=true)\ P(Cavity=true)$

$P(Weather= sunny, Cavity=false) = P(Weather= sunny \mid Cavity= false)\ P(Cavity= false)$
… … … … … … … … … … … …

- Chain rule
- derivata dall'applicazione ripetuta della product rule $P(a \wedge b) = P(b \mid a)\, P(a)$

$$
\begin{aligned}
\mathbf{P}(X_1, \ldots, X_n) &= \mathbf{P}(X_n \mid X_1, \ldots, X_{n-1})\, \mathbf{P}(X_1, \ldots, X_{n-1}) \\
&= \mathbf{P}(X_n \mid X_1, \ldots, X_{n-1})\, \mathbf{P}(X_{n-1} \mid X_1, \ldots, X_{n-2})\, \mathbf{P}(X_1, \ldots, X_{n-2}) \\
&= \ldots \\
&= \prod_i \mathbf{P}(X_i \mid X_1, \ldots, X_{i-1})
\end{aligned}
$$

Esempio di utilizzo delle probabilità condizionate

evidenza

Evidenza ulteriore

Il grado di credenza cresce

P(Batteria= scarica |Auto= non parte) = 0.6

Conoscenza di tipo statistico

P(Batteria=scarica |Auto=non parte $\wedge$ *LasciatiFariAccesi=true)= 0.9*

Evidenza da racconto del proprietario

P(Batteria= scarica |Auto= non parte $\wedge$ *LasciatiFariAccesi=true* $\wedge$ *Fanale=rotto)*
= P(Batteria=scarica |Auto=non parte $\wedge$ *LasciatiFariAccesi=true)=0.9*

Evidenza da osservazione (irrilevante), si semplifica

P(Batteria=scarica |Auto= non parte $\wedge$ *LasciatiFariAccesi=true* $\wedge$
MisuraBatteriaScarica=true)= 1

Evidenza da misura dello stato di carica della batteria

- Semantica: (assiomi)
- Per ogni proposizione *a* e *b*:

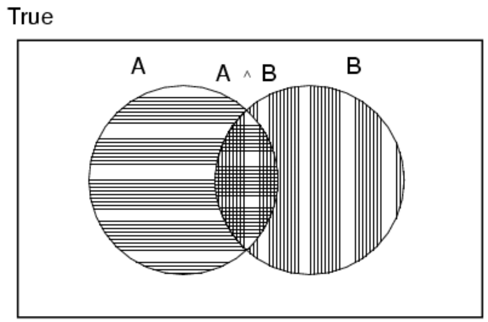
$$0 \leq P(a) \leq 1$$

Tutte le probabilità sono tra 0 e 1

$$P(true) = 1 \qquad P(false) = 0$$

Le proposizioni necessariamente vere (valide) hanno probabilità 1, quelle necessariamente false (non soddisfacibili) hanno probabilità 0

$$P(a \vee b) = P(a) + P(b) - P(a \wedge b)$$



Probabilità di proposizioni legate

da operatori logici

Nota: solo per probabilità a priori. Le probabilità

a posteriori sono state definite nei termini di probabilità a priori

- Dagli assiomi si costruisce il resto della teoria della probabilità
- In particolare si deriva:

La probabilità di una proposizione *a* è la somma delle probabilità di tutti [*e(a)*] gli eventi atomici e_i in cui la proposizione è vera

$$P(a) = \sum_{e_i \in e(a)} P(e_i)$$

Data la full joint distribution, l'equazione fornisce un metodo per calcolare la probabilità di ogni proposizione

- Inferenza probabilistica utilizzando la full joint distribution (inferenza per enumerazione).

Data una qualunque proposizione calcolare la probabilità a priori e a posteriori a partire dalle evidenze disponibili.

Probabilità di un evento atomico

Base di conoscenza

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

Somma delle probabilità = 1

Full joint distribution per il mondo composto dalle variabili *Toothache, Cavity, Catch*
(Mal di denti, Carie, Il ferro del dentista entra nel dente)

Inferenza probabilistica per enumerazione

- La probabilità di una proposizione si calcola identificando gli eventi atomici in cui la proposizione è vera e sommando le relative probabilità

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

P(*toothache*) = 0.108 + 0.012 + 0.016 + 0.064 = 0.2

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

$P(cavity \vee toothache) = .108 + .012 + .072 + .008 + .016 + .064 = 0.28$

Inferenza probabilistica per enumerazione

- Si possono anche calcolare le probabilità condizionate (nei termini di probabilità incondizionate)

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)}$$

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

$$\mathrm{P}(\neg cavity \mid toothache) = \frac{\mathrm{P}(\neg cavity \wedge toothache)}{\mathrm{P}(toothache)}$$

$$= \frac{0.016+0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

- Normalizzazione

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = \frac{0.016+0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)} = \frac{0.108+0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

Il denominatore non cambia. $\frac{1}{P(toothache)}$ Può essere visto come costante di normalizzazione per la distribuzione $\mathbf{P}(\underset{\uparrow}{Cavity} \mid \underset{\uparrow}{toothache})$ (garantisce che la somma sia 1)

Variabile *Toothache=true*

Inferenza probabilistica per enumerazione

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

Joint probability distribution di *Cavity e toothache*
Si calcola con

$$P(a) = \sum_{e_i \in e(a)} P(e_i)$$

La coppia delle due precedenti equazioni può essere scritta come:

$\mathbf{P}(Cavity \mid toothache) = \alpha\, \mathbf{P}(Cavity, toothache)$

variabile

$= \alpha\, [\mathbf{P}(Cavity, toothache, catch) + \mathbf{P}(Cavity, toothache, \neg\, catch)]$

$= \alpha\, [<0.108, 0.016> + <0.012, 0.064>]$

$= \alpha\, <0.12, 0.08> = <0.6, 0.4>$

$\alpha = 1/(0.108 + 0.012 + 0.016 + 0.064) = 1/\ 0{,}2$

Idea generale: calcolare la distribuzione della *query variable* fissando le *evidence variables* e sommando le probabilità delle *hidden variables*

Inferenza probabilistica per enumerazione

- Notazione:

Y Hidden variables
rimanenti variabili non osservate

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

$$P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)} = \frac{0.016+0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

E Evidence variables
variabili osservate

X The query variable

Inferenza probabilistica per enumerazione

- Procedura di inferenza per rispondere a query probabilistiche per variabili discrete (query su singola variabile) (per enumerazione delle entries in una full joint distribution) Come calcolare una probabilità condizionata
- X Query variable, **E** Evidence variables, **e** valori osservati per **E**, **Y** Hidden variables i cui valori sono **y**

Query: $\mathbf{P}(X \mid \mathbf{e})$

Distribuzione di probabilità di X condizionata da **e**

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X,\mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X,\mathbf{e},\mathbf{y})$$

$$\alpha = \frac{1}{\mathbf{P}(\mathbf{e})}$$

Somma su tutti i possibili valori **y** delle variabili non osservate

- Problemi
  - Tempo: complessità nel caso peggiore $O(d^n)$ dove d è la più grande arità delle n variabili
  - Memoria: $O(d^n)$ per memorizzare la full joint distribution
  - Trovare i numeri della full joint distribution. Esperienza richiesta per stimare ogni elemento della tabella. Impraticabile su problemi realistici

	toothache		¬ *toothache*	
	catch	¬ *catch*	*catch*	¬ *catch*
cavity	.108	.012	.072	.008
¬ *cavity*	.016	.064	.144	.576

- Indipendenza assoluta
- Un modo per ridurre l’informazione necessaria per specificare la full joint distribution
- Introduciamo un’altra variabile

Weather <sunny,rainy,cloudy,snow>

La tabella della full joint distribution ha 32 (8x4) elementi da definire

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

Una tabella per ogni valore della variabile *Weather*

- Si suppone che il tempo atmosferico non sia influenzato dai problemi dentali

P(*Weather= cloudy* | *toothache, catch, cavity*) = P(*Weather= cloudy)*

Indipendenza assoluta

- In generale *a* e *b* sono indipendenti iff
  $P(a|b) = P(a)$ or $P(b|a) = P(b)$ or

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} = P(a)$$

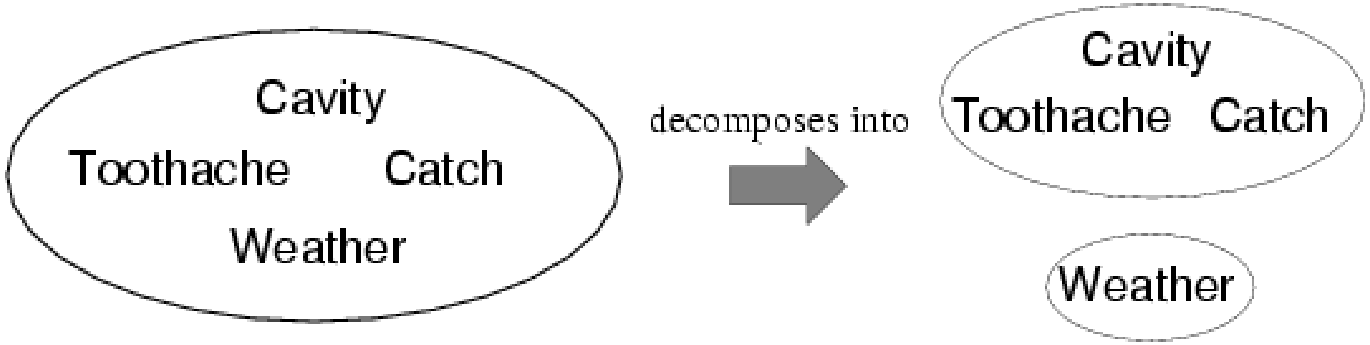
$$\downarrow$$

$$\boxed{P(a \wedge b) = P(a)\, P(b)}$$

- In questo caso si può scrivere:

P(*Toothache, Catch, Cavity, Weather*) =
P(*Toothache, Catch, Cavity*) **P**(*Weather*)

Ci si è ridotti da una tabella di 32 elementi a due tabelle, una di 8 elementi e l'altra di 4 elementi (12 elementi)



	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

	toothache		¬ toothache	
	catch	¬ catch	catch	¬ catch
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

sunny	0,4
rainy	0,2
cloudy	0,3
snow	0,1

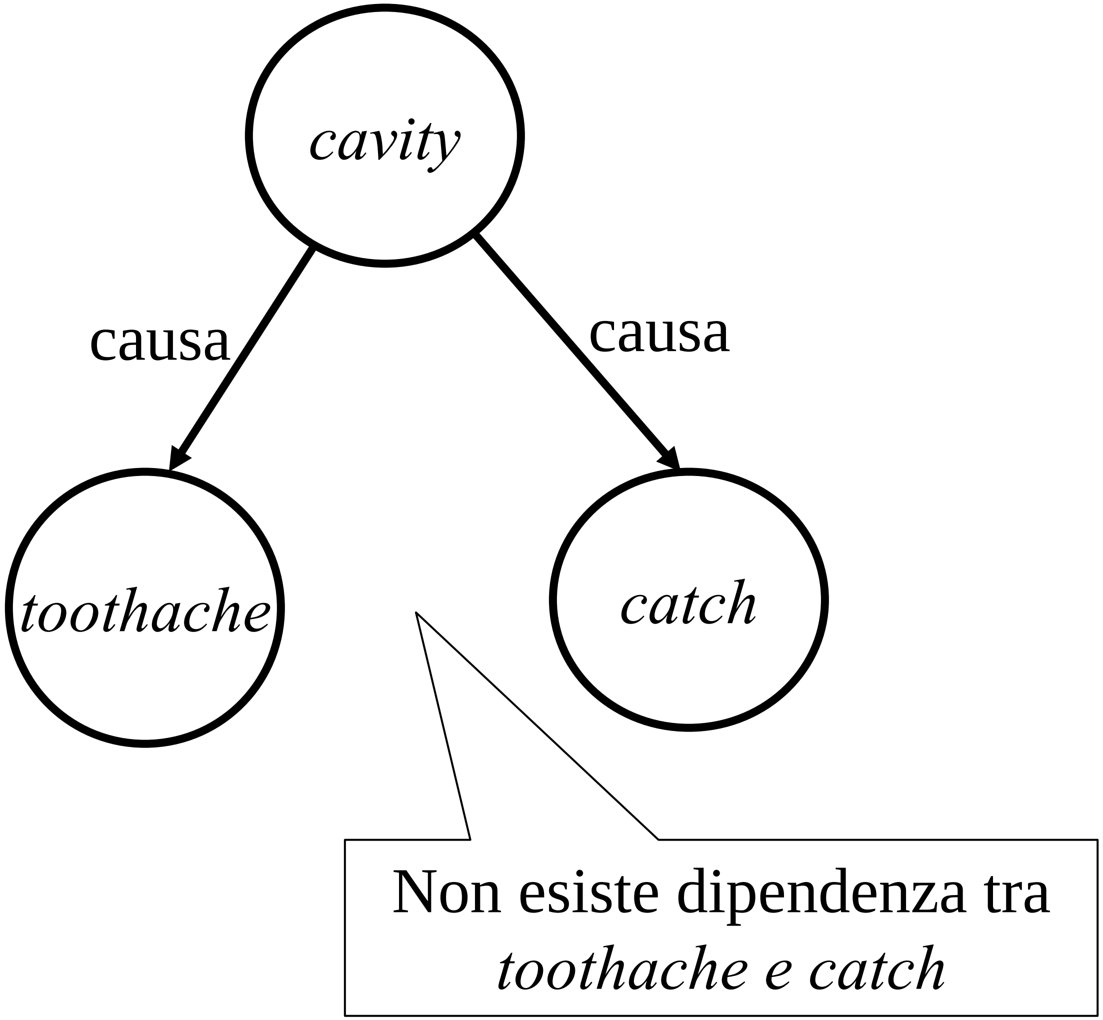
- L’indipendenza assoluta è basata su conoscenza del dominio
- E’ rara
- Le singole tabelle possono essere comunque grandi

- Indipendenza condizionale
- **P**(*Toothache, Cavity, Catch*) ha $2^3 - 1 = 7$ elementi indipendenti

- Se ho *cavity*, la probabilità di *catch* non dipende da *toothache*:
  P(*catch* | *toothache, cavity*) = P(*catch* | *cavity*)
- Lo stesso vale se non ho cavity:
  P(*catch* | *toothache,*¬*cavity*) = P(*catch* | ¬*cavity*)

- *Catch* è condizionalmente indipendente da *Toothache,* data *Cavity*:
  P(*Catch* | *Toothache,Cavity*) = **P**(*Catch* | *Cavity*)
- Equivalentemente
  P(*Toothache* | *Catch, Cavity*) = **P**(*Toothache* | *Cavity*)
  P(*Toothache, Catch* | *Cavity*) = **P**(*Toothache* | *Cavity*) **P**(*Catch* | *Cavity*)

Indipendenza



- Come per l'indipendenza assoluta si può scrivere (usando la chain rule)

Joint probability distribution

$P(a \wedge b) = P(a \mid b)\, P(b)$ (product rule)
+ relazione di indipendenza condizionale

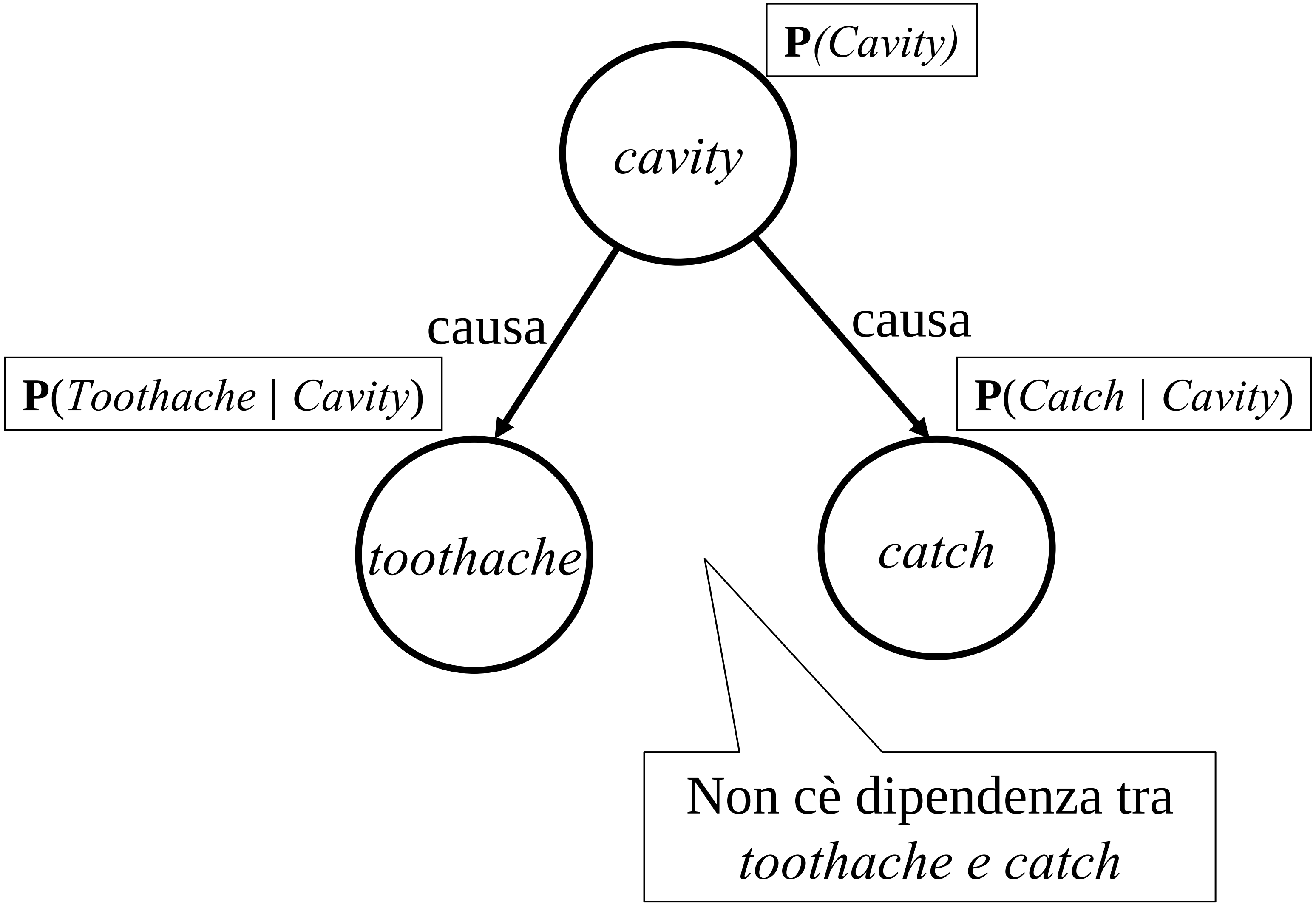
$\mathbf{P}(Toothache, Catch, Cavity)$
$= \mathbf{P}(Toothache \mid Catch, Cavity)\, \mathbf{P}(Catch, Cavity)$
$= \mathbf{P}(Toothache \mid Catch, Cavity)\, \mathbf{P}(Catch \mid Cavity)\, \mathbf{P}(Cavity)$

Usando la relazione di indipendenza:
$\mathbf{P}(Toothache \mid Catch, Cavity) = \mathbf{P}(Toothache \mid Cavity)$

$= \mathbf{P}(Toothache \mid Cavity)\, \mathbf{P}(Catch \mid Cavity)\, \mathbf{P}(Cavity)$

- La tabella originale è decomposta in tre tabelle più piccole

Indipendenza



Indipendenza

$$\mathbf{P}(Toothache \mid Cavity)\ \mathbf{P}(Catch \mid Cavity)\ \mathbf{P}(Cavity)$$

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	.108	.012	.072	.008
¬ cavity	.016	.064	.144	.576

7 numeri indipendenti
(somma probabilità = 1)

	P(toothache| Cavity)	*P(¬toothache| Cavity)*
cavity		
¬ cavity		

2 numeri indipendenti

$$P(A \mid B) + P(\neg A \mid B) = 1$$

	P(catch| Cavity)	*P(¬catch| Cavity)*
cavity		
¬ cavity		

2 numeri indipendenti

	P(Cavity)
cavity	
¬ cavity	

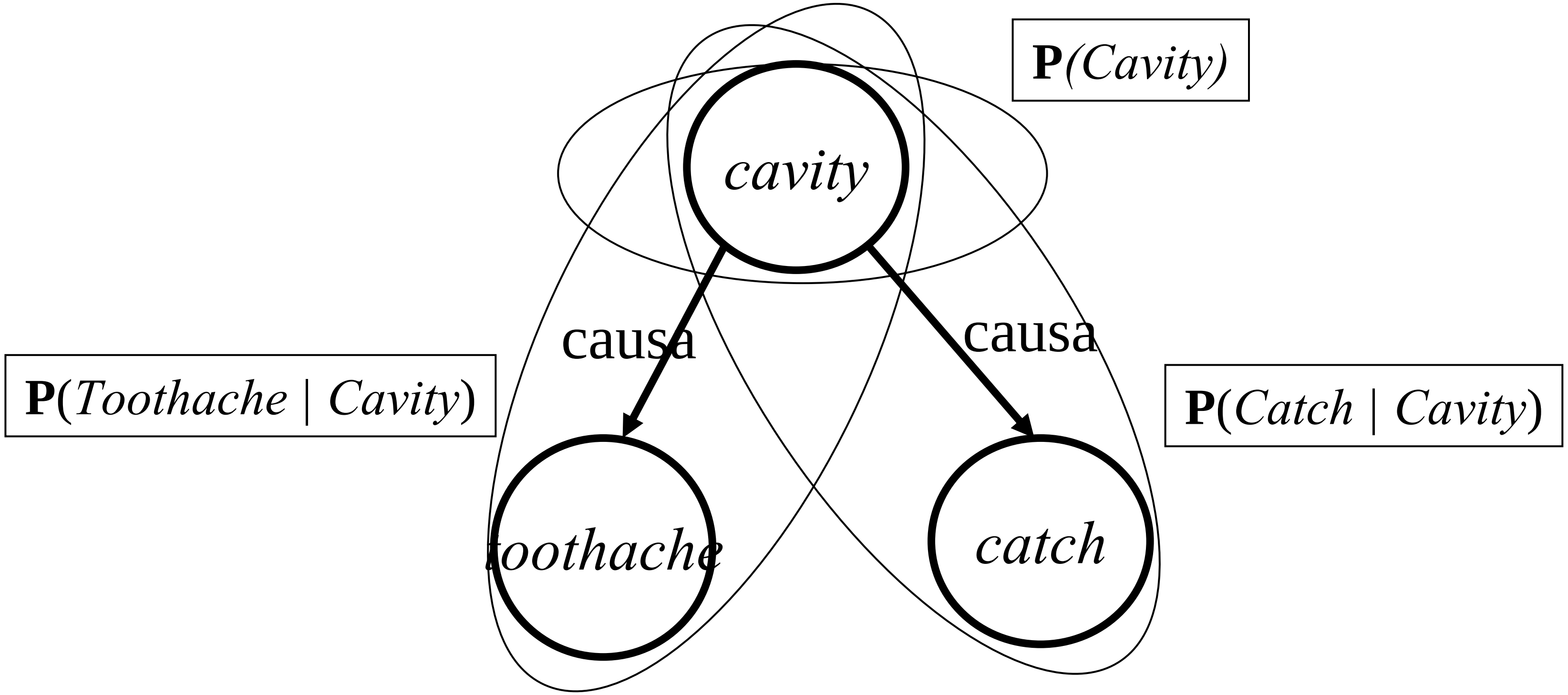
1 numero indipendente

Totale: 5 numeri

Indipendenza

- La tabella originale è decomposta in tre tabelle più piccole
  - Memoria: passa da *$O(d^n)$* a *$O(n)$ ove d* è la più grande arità delle n variabili
  - L'indipendenza condizionata è basata su conoscenza del dominio
  - L'indipendenza condizionata è più comune

Indipendenza



L'indipendenza condizionale permette la separazione di grandi domini probabilistici in sottoinsiemi debolmente connessi
La full joint distribution può essere scritta:

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause) \prod_i \mathbf{P}(Effect_i \mid Cause)$$

- Product rule $P(a \wedge b) = P(a \mid b)\, P(b)$

  $P(a \wedge b) = P(b \mid a)\, P(a)$

dividendo per $P(b)$:

$$P(a \mid b) = \frac{P(b \mid a)\, P(a)}{P(b)}$$

Regola di Bayes

Bayes, Thomas (1763) An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, **53:370-418**

Direzione causale

Direzione diagnostica

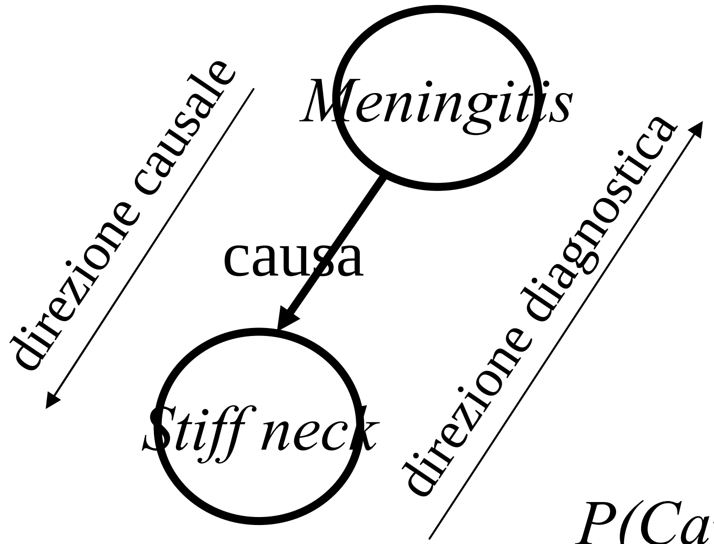
$$P(Cause \mid Effect) = \frac{P(Effect \mid Cause)\, P(Cause)}{P(Effect)}$$

- Caso più generale con variabili a più valori:

$$\boldsymbol{P}(Y \mid X) = \frac{\boldsymbol{P}(X \mid Y)\,\boldsymbol{P}(Y)}{\boldsymbol{P}(X)}$$

- La regola è la base dei moderni sistemi di inferenza probabilistica

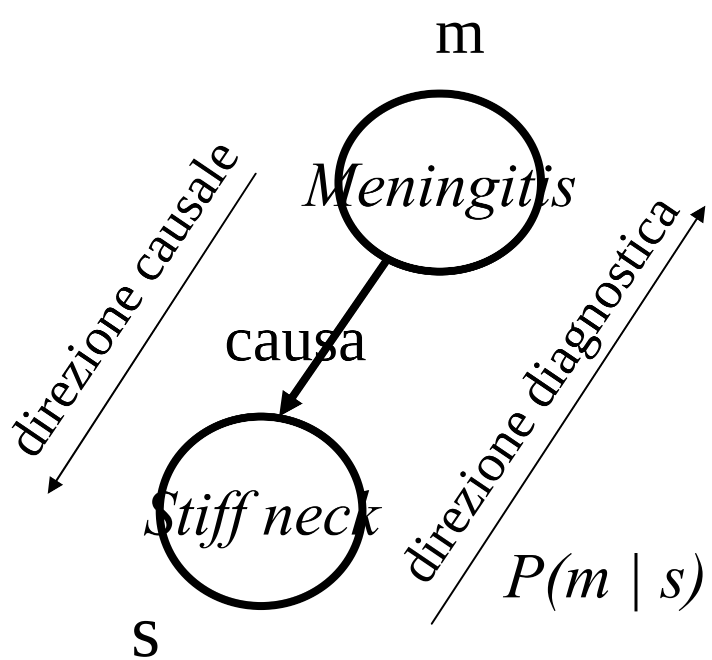
- Perchè è utile?
- Calcola una probabilità condizionata a partire da una probabilità condizionata (nella direzione opposta) e due probabilità non condizionate)
- In molti casi si possono stimare i tre numeri e serve calcolare il quarto



$$P(a \mid b) = \frac{P(b \mid a)\, P(a)}{P(b)}$$

$$P(Cause|Effect) = \frac{P(Effect|Cause)\, P(Cause)}{P(Effect)}$$

- Un paziente su 5000 con il collo rigido ha la meningite
- La meningite in molti casi (50%) causa il collo rigido, ma la probabilità in direzione diagnostica è bassa poichè la probabilità a priori del collo rigido (1/20) è molto più alta di quella a priori della meningite (1/50000) (tanta gente ha il collo rigido non causato da meningite)



$P(s \mid m) = 0,5$ *direzione causale*

$P(s) = 1/20$

$P(m) = 1/50000$

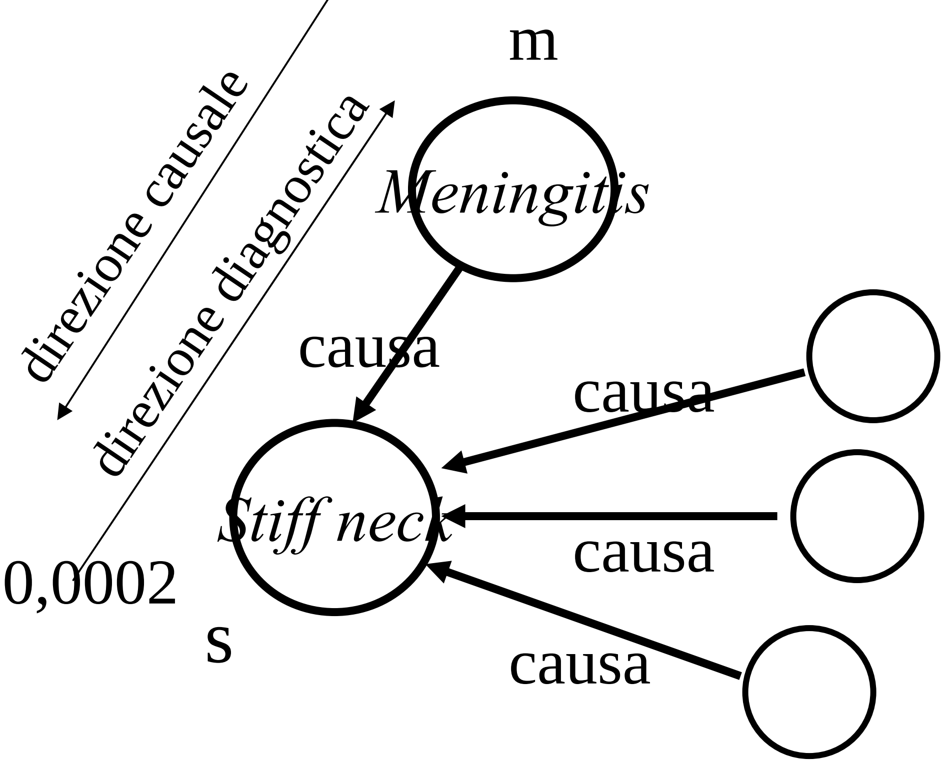
direzione diagnostica

$$P(m \mid s) = \frac{P(s \mid m)\, P(m)}{P(s)} = \frac{0,5 * 1/50000}{1/20} = 0,0002$$

Regola di Bayes

- (tanta gente ha il collo rigido non causato da meningite)

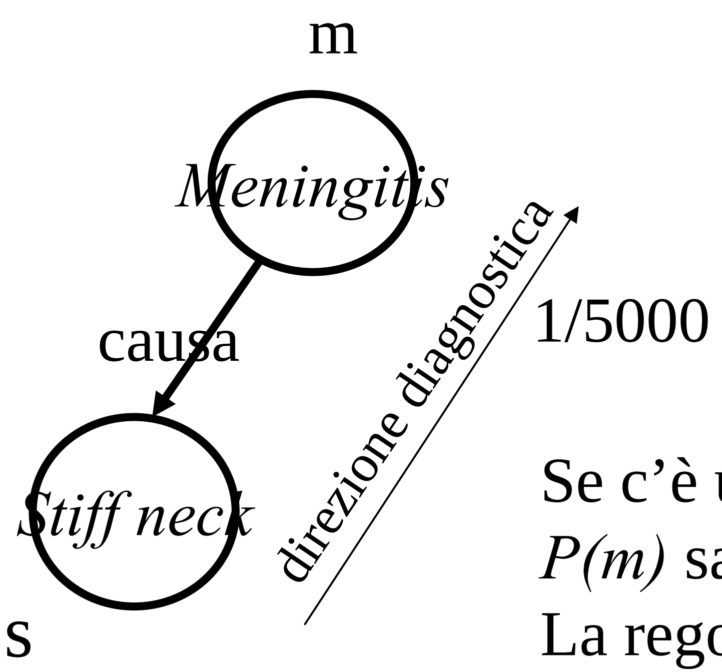
$P(s \mid m) = 0{,}5$



$P(m \mid s) = 0{,}0002$

Regola di Bayes

- Sistema basato su regole diagnostiche (dai sintomi alle cause)
- Da osservazioni di pazienti di ricava che un paziente su 5000 con il collo rigido ha la meningite



KB (Shallow knowledge)

$P(m \mid s) = 0{,}0002$

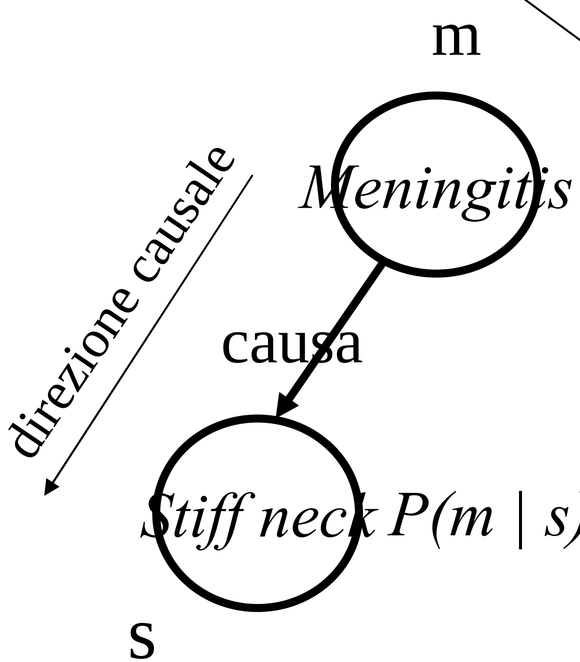
Se c’è una epidemia di meningite
P(m) sale ad es a 1/20000
La regola diagnostica non funziona più
Devono ancora essere accumulate le osservazioni sui pazienti nella nuova situazione

- Sistema basato su regole causali (dalle cause ai sintomi)

Dati statistici su diffusione meningite e diffusione collo rigido

KB (Model based knowledge)

Osservazioni sul numero di pazienti con meningite che hanno il collo rigido



$P(s \mid m) = 0{,}5$ *direzione causale*

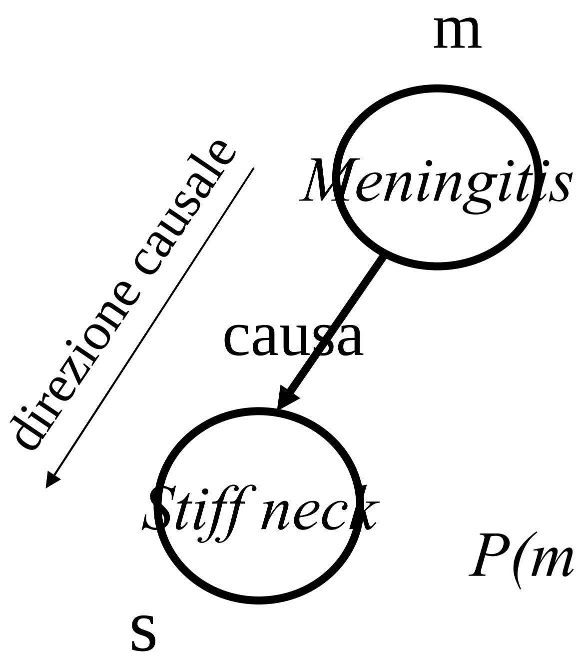
$P(s) = 1/20$

$P(m) = 1/50000$

Inferenza con il teorema di Bayes

$$P(m \mid s) = \frac{P(s \mid m)\,P(m)}{P(s)} = \frac{0{,}5 * 1/50000}{1/20} = 0{,}0002$$

- Sistema basato su regole causali (dalle cause ai sintomi)



KB (Model based knowledge)

$P(s \mid m) = 0,5$ *direzione causale*

$P(s) = 1/20$

$P(m) = \mathbf{1/20000}$

Inferenza con il teorema di Bayes

$$P(m \mid s) = \frac{P(s \mid m)\, P(m)}{P(s)} = \frac{0,5 * \mathbf{1/20000}}{1/20} = \mathbf{0,0005}$$

Si osserva il numero accresciuto di casi di meningite

Il sistema diagnostico funziona ancora

La regola causale è più robusta della regola diagnostica (non è modificata dall’epidemia, esprime un modello di funzionamento della meningite)

- Nota: al posto di assegnare la probabilità a priori dell'evidenza *P(s),* si può calcolare la probabilità a posteriori di ogni valore della *query variable* (m e $\neg m$) e normalizzare il risultato

$$P(m \mid s) = \frac{P(s \mid m)\, P(m)}{P(s)}$$

Vedi normalizzazione, slide 31

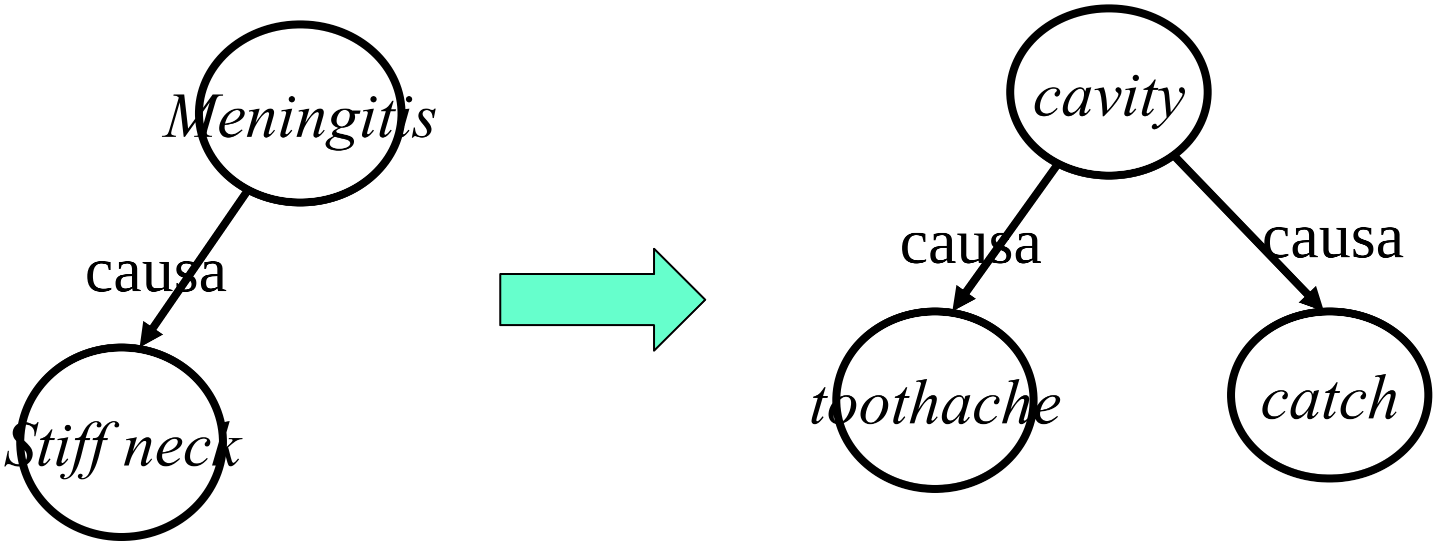
$$\mathbf{P}(M \mid s) = \alpha \langle P(s \mid m)\, P(m),\, P(s \mid \neg m)\, P(\neg m) \rangle$$

Regola di Bayes con normalizzazione

$$\mathbf{P}(Y \mid X) = \alpha\, \mathbf{P}(X \mid Y)\, \mathbf{P}(Y)$$

α Costante necessaria per rendere
la somma degli elementi di $\mathbf{P}(Y \mid X) = 1$

- Combinare le evidenze



- Con la full joint distribution possiamo calcolare ad esempio (problemi di scalabilità- $O(d^n)$):

$$\mathbf{P}(Cavity \mid toothache \wedge catch)$$

- Con il teorema di Bayes:

$$\mathbf{P}(Cavity \mid toothache \wedge catch) =$$

$$\alpha \mathbf{P}(toothache \wedge catch \mid Cavity)\, \mathbf{P}(Cavity)$$

dobbiamo conoscere le probabilità condizionali della congiunzione t*oothache* $\wedge$ *catch* per ogni valore di *Cavity*

2^n se n=numero evidenze - problema di scalabilità

- Utilizzo della nozione di indipendenza condizionale

$$\mathbf{P}(Cause, Effect_1, \ldots, Effect_n) = \mathbf{P}(Cause) \prod_i \mathbf{P}(Effect_i \mid Cause)$$

Si risponde ad una query diagnostica usando tre tabelle di probabilità:
•una a priori
•due condizionate in verso causale

Direzione diagnostica

$$\mathbf{P}(Cavity \mid toothache \wedge catch)$$

$$= \alpha\, \mathbf{P}(toothache \wedge catch \mid Cavity)\, \mathbf{P}(Cavity)$$

Usando la relazione di indipendenza (slide 39):

$$\mathbf{P}(Toothache, Catch \mid Cavity) = \mathbf{P}(Toothache \mid Cavity)\mathbf{P}(Catch \mid Cavity)$$

Direzione causale

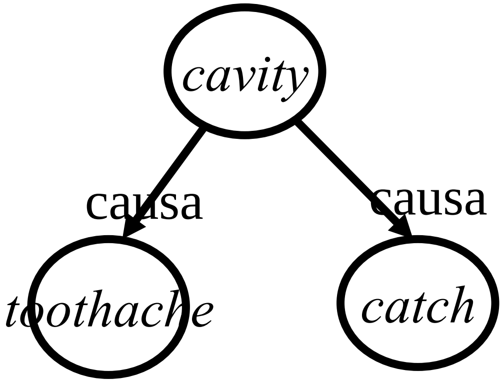
$$= \alpha\, \mathbf{P}(toothache \mid Cavity)\, \mathbf{P}(catch \mid Cavity)\, \mathbf{P}(Cavity)$$

Regola di Bayes

$\mathbf{P}(Cavity \mid toothache \wedge catch) = \alpha\, \mathbf{P}(toothache \mid Cavity)\, \mathbf{P}(catch \mid Cavity)\, \mathbf{P}(Cavity)$

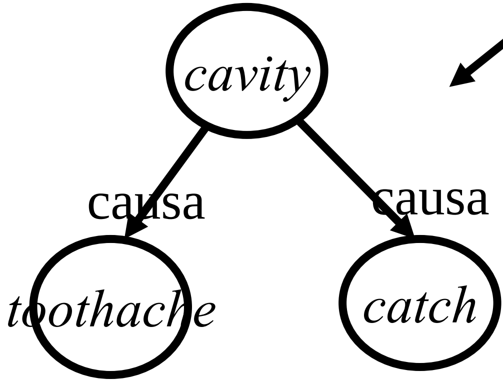
$\mathbf{P}(Cavity \mid toothache \wedge catch)$

Verso diagnostico



Verso causale (teorema di Bayes)

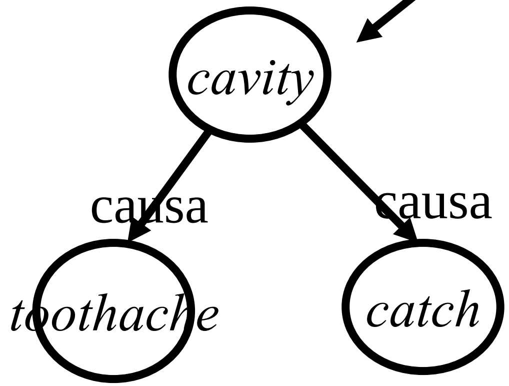
$\alpha\, \mathbf{P}(Cavity)$



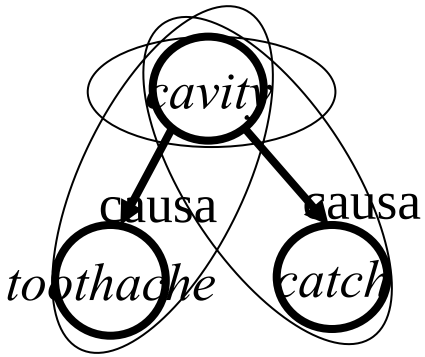
indipendenza condizionale

$\mathbf{P}(toothache \wedge catch \mid Cavity)$

$\alpha\, \mathbf{P}(Cavity)$



$\mathbf{P}(toothache \mid Cavity)\quad \mathbf{P}(catch \mid Cavity)$



- E’ un esempio di Naive Bayes model (Bayesian classifier)
- Si assume che una singola causa influenza direttamente più effetti, tutti condizionalmente indipendenti, data la causa.
- Usato come semplificazione anche quando non è vera l’indipendenza condizionale

Sintesi

- Sintesi:
  - Il trattamento della conoscenza incerta è necessario per la costruzione di molti sistemi reali
  - La teoria della probabilità fornisce lo strumento formale per riassumere le credenze di un agente
  - La probabilità condizionata descrive le variazioni di credenze al variare delle evidenze disponibili
  - …

Sintesi

- …
- Una tavola completa di distribuzione di probabilità (full joint probability distribution) può essere utilizzata per inferire (con risorse esponenziali) la probabilità di una proposizione
- L’indipendenza assoluta e l’indipendenza condizionale tra sottoinsiemi di variabili fattorizzano la tavola riducendo la complessità
- …

Sintesi

- …
- La regola di Bayes permette di calcolare probabilità condizionate non note (es. in direzione diagnostica su una rete causale) a partire da probabilità condizionate (in direzione causale) e probabilità a priori
- La complessità di calcolo della regola di Bayes può essere ridotta utilizzando le indipendenze condizionali