Reti Bayesiane

Corso di Sistemi Intelligenti, Capitolo 14.1–3

Outline

- Sintassi
- Semantica
- Distribuzioni parametrizzate

Reti Bayesiane (Bayesian networks)

Una semplice notazione grafica per asserzioni condizionalmente indipendenti e quindi per specifiche di distribuzioni condizionali complete Sintassi:

un insieme di nodi, uno per variabile

un grafo diretto aciclico (link $\approx$ "influenza direttamente")

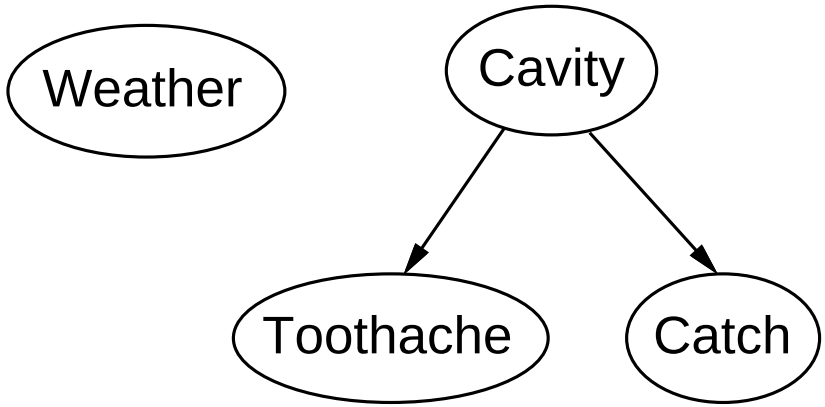
una distribuzione condizionale per ogni nodo dati i suoi genitori:

$\mathbf{P}(X_i|Parents(X_i))$

Nel caso più semplice, distribuzione condizionale rappresentata come una tabella della probabilità condizionale (CPT) data la distribuzione su X_i per ogni combinazione di valori assunti dai genitori

Esempio

La topologia della rete codifica asserzioni di indipendenza condizionale:



$Weather$ è indipendente dalle altre variabili

$Toothache$ e $Catch$ sono condizionalmente indipendenti data $Cavity$

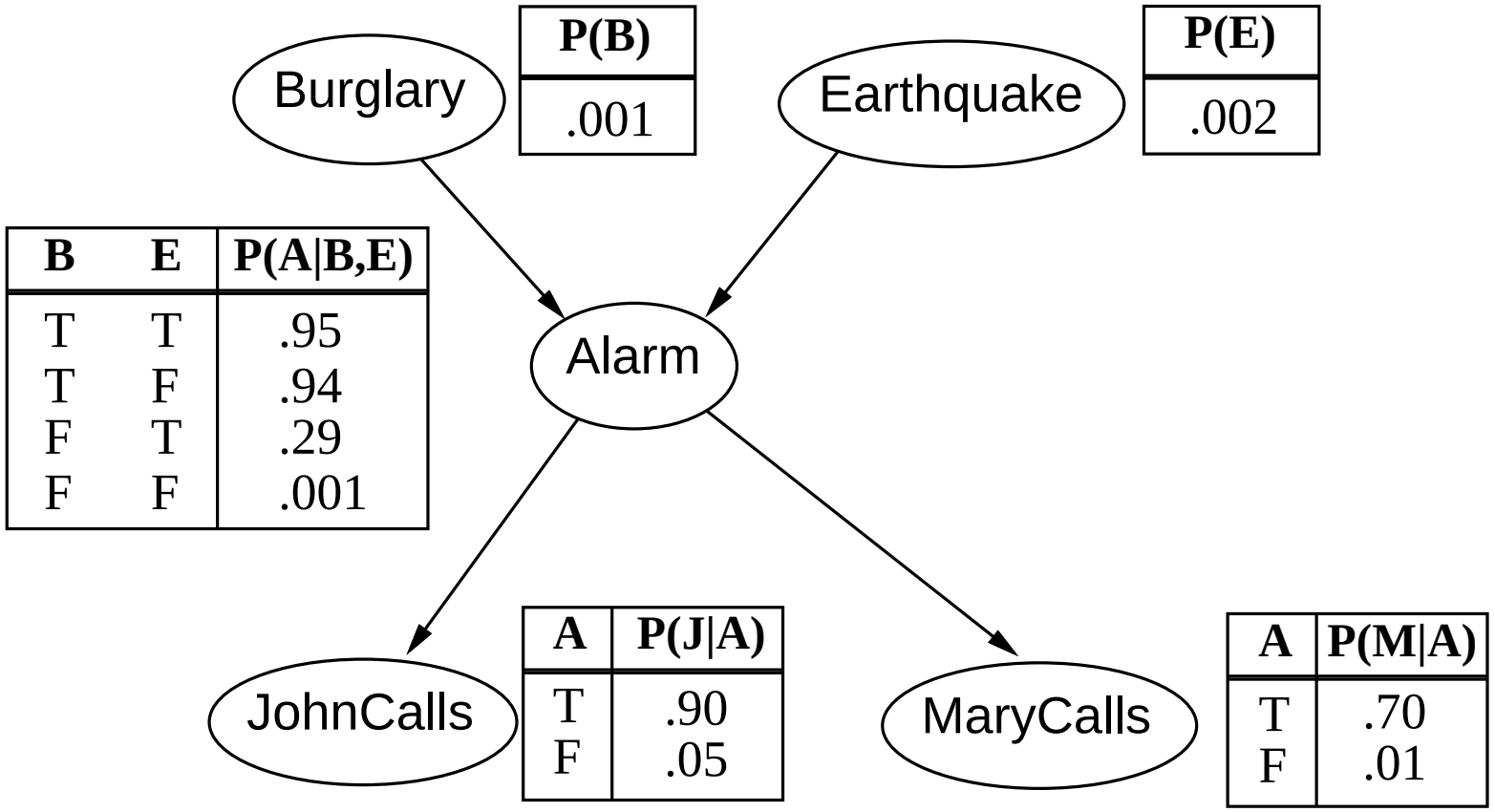
Esempio

Sono al lavoro, il vicino John chiama per dire che il mio allarme $Alarm$ è entrato in funzione, ma la vicina Mary non chiama. Alcune volte l'allarme è attivato da piccole scosse di terremoto. C'è un ladro in casa ?

Variabili: $Burglar$, $Earthquake$, $Alarm$, $JohnCalls$, $MaryCalls$
La topologia della rete riflette conoscenza "causale":

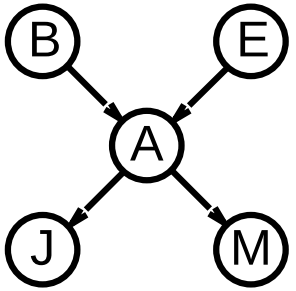
- Un ladro può attivare l'allarme
- Un terremoto può attivare l'allarme
- L'attivazione dell'allarme può indurre Mary a chiamare
- L'attivazione dell'allarme può indurre John a chiamare

Esempio


Burglary
P(B)
.001
Earthquake
P(E)
.002
B E P(A|B,E)
T T .95
T F .94
F T .29
F F .001
Alarm
JohnCalls
A P(J|A)
T .90
F .05
MaryCalls
A P(M|A)
T .70
F .01

Compattezza

Una CPT per variabili Booleane X_i con k genitori Booleani ha 2^k righe per le combinazioni di valori dei genitori



Ogni riga richiede un numero p per $X_i = vero$
(il numero per $X_i = falso$ è $1 - p$)

Se ogni variabile non ha più di k genitori,
la rete completa richiede $O(n \cdot 2^k)$ numeri

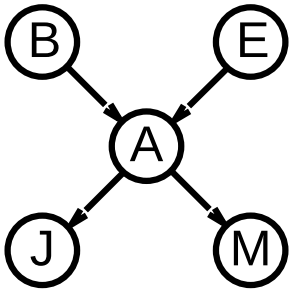
Cioè, cresce linearmente con n, vs. $O(2^n)$ per la distribuzione congiunta completa

Per la rete precedente, $1 + 1 + 4 + 2 + 2 = 10$ numeri (vs. $2^5 - 1 = 31$)

Semantica globale

La semantica globale definisce la distribuzione congiunta completa come il prodotto delle distribuzioni condizionali locali:

$$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i))$$



p.e., $P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$=$

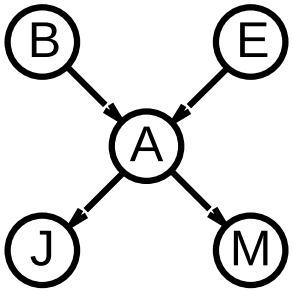
Semantica globale

La semantica globale definisce la distribuzione congiunta completa come il prodotto delle distribuzioni condizionali locali:

$$\mathbf{P}(X_1, \ldots, X_n) = \prod_{i=1}^{n} \mathbf{P}(X_i | Parents(X_i))$$

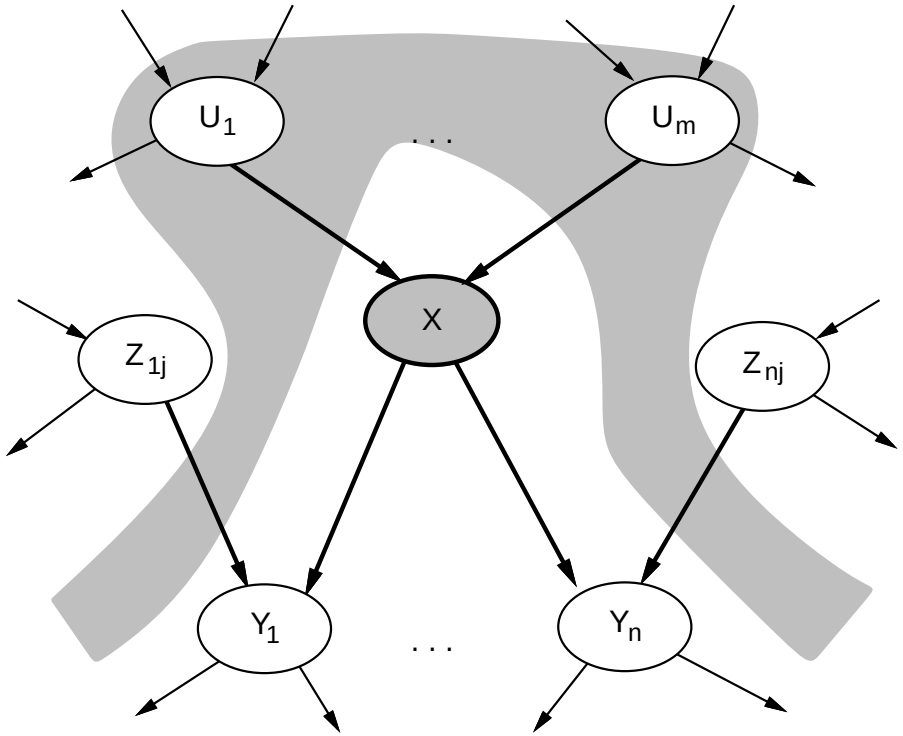
p.e., $P(j \wedge m \wedge a \wedge \neg b \wedge \neg e)$

$$= P(j|a)P(m|a)P(a|\neg b, \neg e)P(\neg b)P(\neg e)$$



Semantica locale

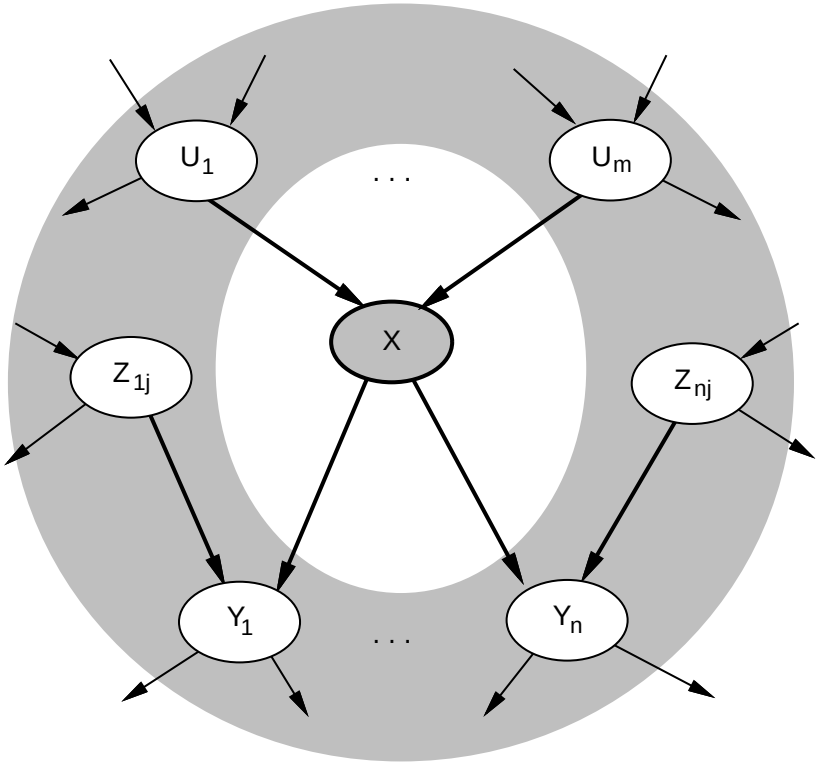
Semantica locale: ogni nodo è condizionalmente indipendente dai suoi non discendenti dati i genitori



Teorema: Semantica locale $\Leftrightarrow$ semantica globale

Markov blanket

Ogni nodo è condizionalmente indipendente da tutti gli altri dato il suo Markov blanket: genitori + figli + genitori dei figli



Construzione di Reti Bayesiane

Necessità di un metodo tale che data una serie di asserzioni di indipendenza condizionale localmente controllabili, garantisca la semantica globale desiderata

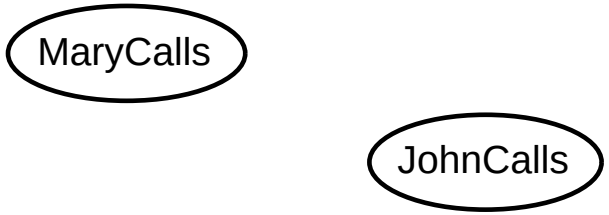
1. Scegliere un ordinamento di variabili $X_1, \ldots, X_n$
2. For $i = 1$ to n
   aggiungi X_i alla rete
   seleziona genitori da $X_1, \ldots, X_{i-1}$ tali che
   $\mathbf{P}(X_i|Parents(X_i)) = \mathbf{P}(X_i|X_1, \ldots, X_{i-1})$

Questa scelta di genitori garantisce la semantica globale:

$$\begin{aligned}\mathbf{P}(X_1, \ldots, X_n) &= \prod_{i=1}^{n} \mathbf{P}(X_i|X_1, \ldots, X_{i-1}) \quad \text{(chain rule)} \\ &= \prod_{i=1}^{n} \mathbf{P}(X_i|Parents(X_i)) \quad \text{(per costruzione)}\end{aligned}$$

Esempio

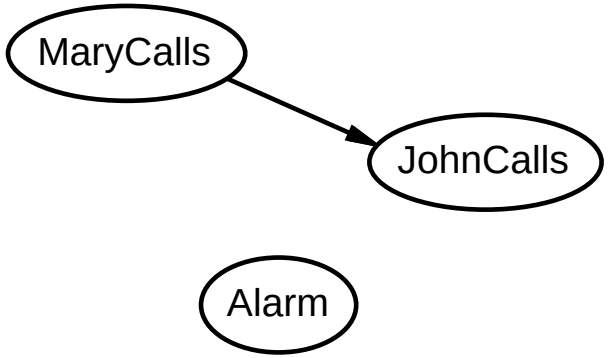
Supponiamo di scegliere l'ordine M, J, A, B, E



$P(J|M) = P(J)$?

Esempio

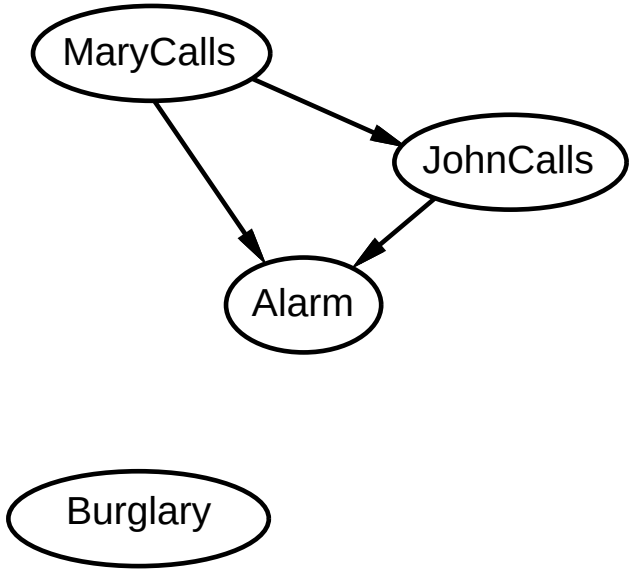
Supponiamo di scegliere l'ordine M, J, A, B, E



$P(J|M) = P(J)$? No
$P(A|J, M) = P(A|J)$? $P(A|J, M) = P(A)$?

Esempio

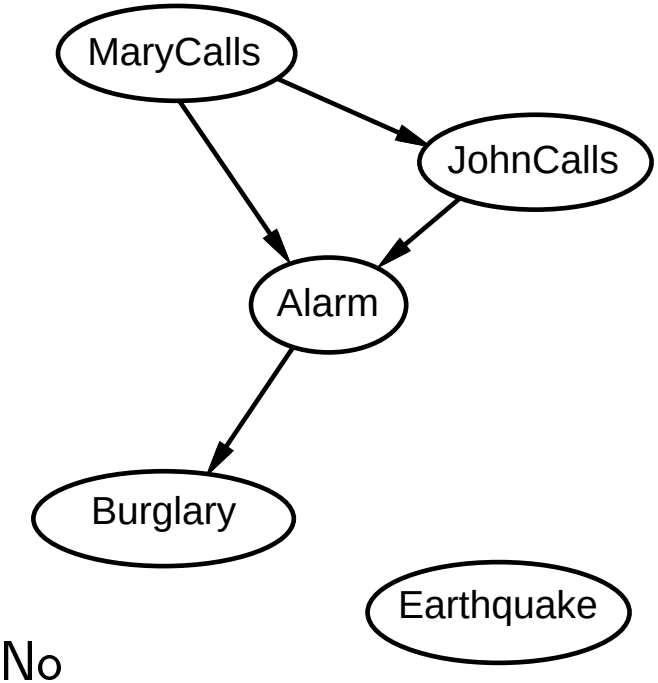
Supponiamo di scegliere l'ordine M, J, A, B, E



$P(J|M) = P(J)$? No
$P(A|J, M) = P(A|J)$? $P(A|J, M) = P(A)$? No
$P(B|A, J, M) = P(B|A)$?
$P(B|A, J, M) = P(B)$?

Esempio

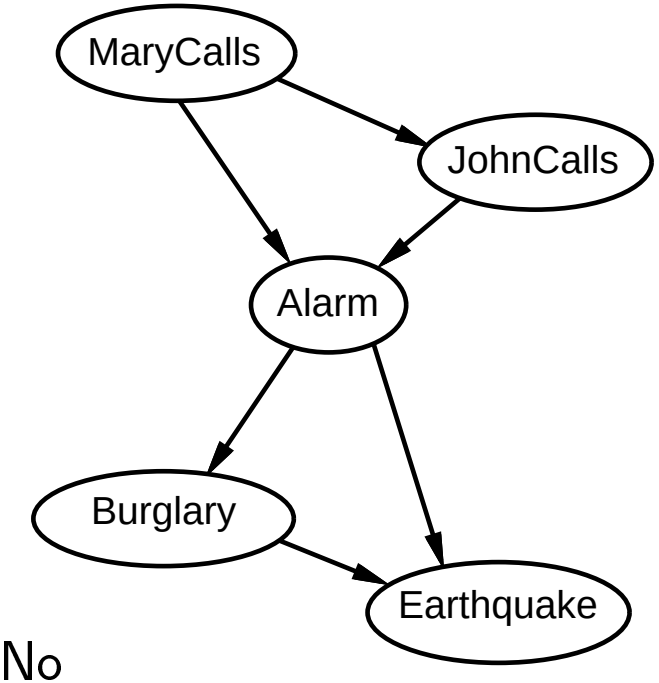
Supponiamo di scegliere l'ordine M, J, A, B, E



$P(J|M) = P(J)$? No
$P(A|J, M) = P(A|J)$? $P(A|J, M) = P(A)$? No
$P(B|A, J, M) = P(B|A)$? Yes
$P(B|A, J, M) = P(B)$? No
$P(E|B, A, J, M) = P(E|A)$?
$P(E|B, A, J, M) = P(E|A, B)$?

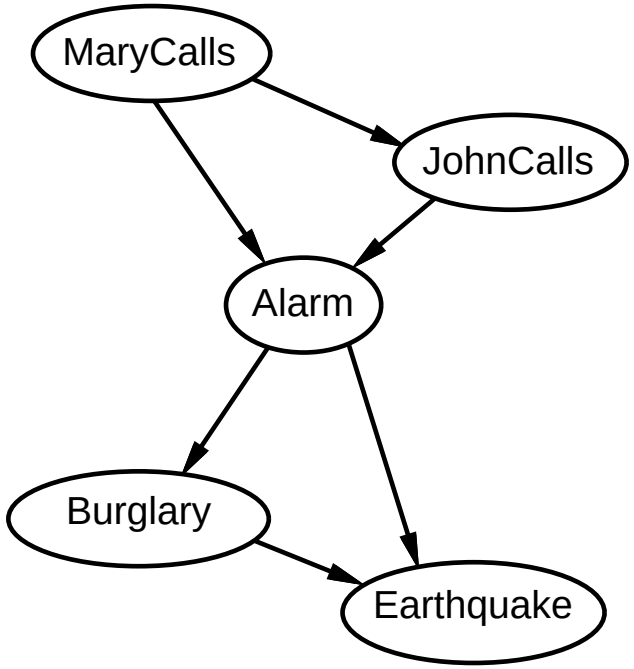
Esempio

Supponiamo di scegliere l'ordine M, J, A, B, E



$P(J|M) = P(J)$? No
$P(A|J,M) = P(A|J)$? $P(A|J,M) = P(A)$? No
$P(B|A,J,M) = P(B|A)$? Yes
$P(B|A,J,M) = P(B)$? No
$P(E|B,A,J,M) = P(E|A)$? No
$P(E|B,A,J,M) = P(E|A,B)$? Yes

Esempio



Decidere l'indipendenza condizionale è difficile nelle direzioni non causali

Valutare le probabilità condizionali è difficile in direzioni non causali

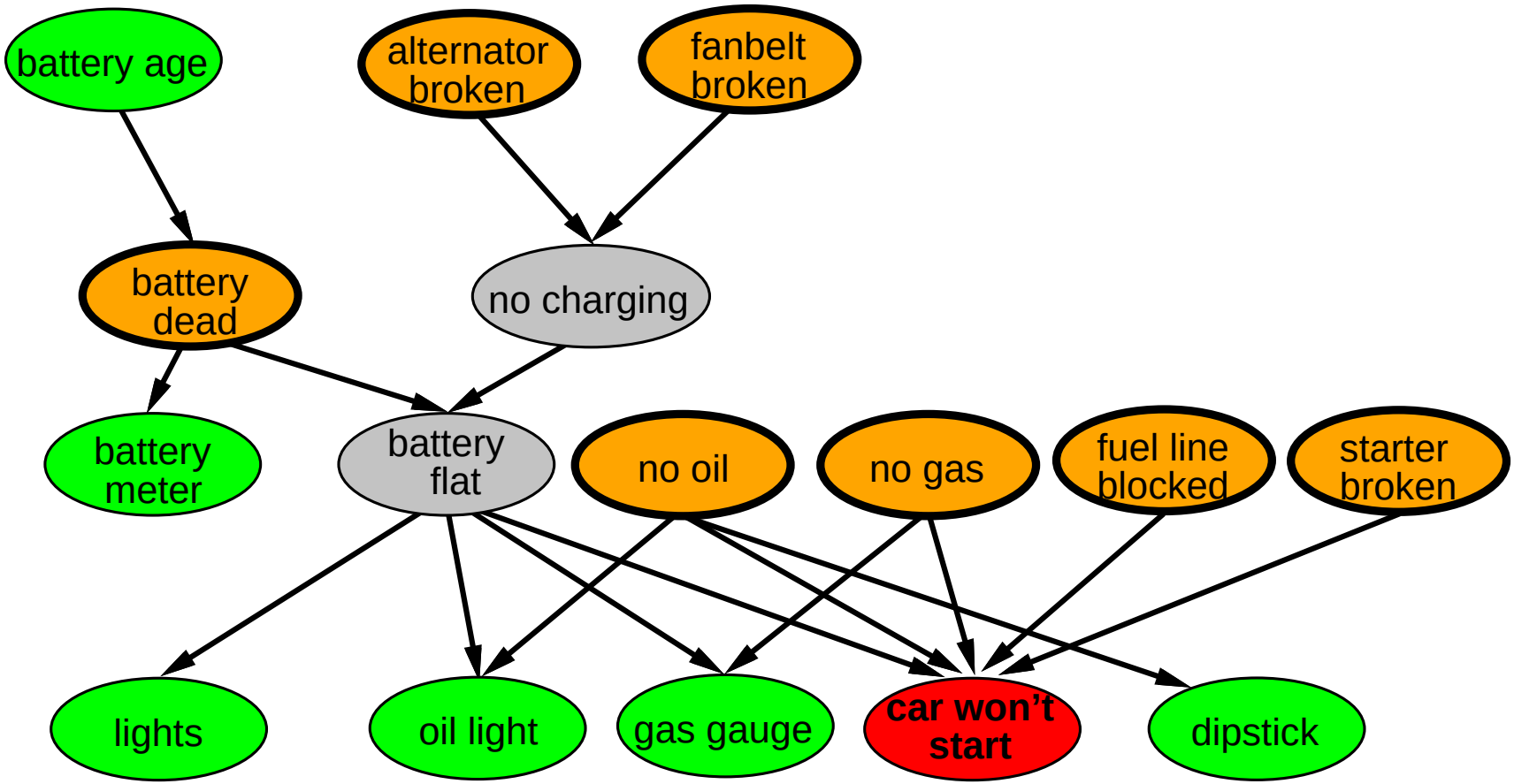
La rete è meno compatta: $1+2+4+2+4=13$ numeri necessari

Esempio: diagnosi per automobile

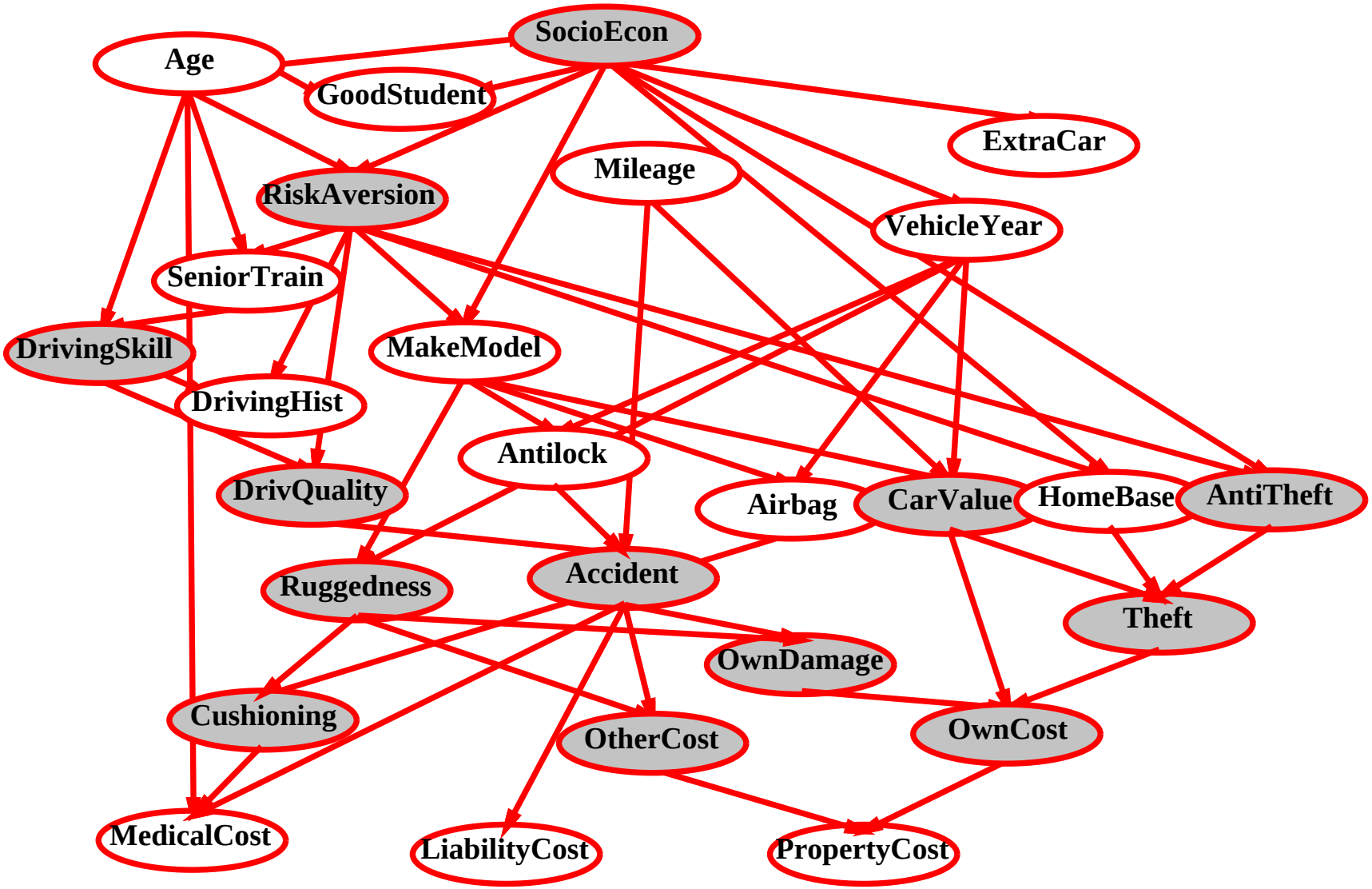
Evidenza iniziale: auto non parte

Variabili controllabili (in verde), variabili "rotto, da aggiustare" (in arancio)

Variabili nascoste (in grigio) assicurano struttura sparsa, riducono i parametri



Esempio: assicurazione dell'automobile



Distribuzioni condizionali compatte

CPT cresce esponenzialmente con il n. di genitori
CPT diventa infinita con genitori o figli a valori continui

Soluzione: distribuzioni canoniche che sono definite in modo compatto

I nodi deterministici sono il caso più semplice:
$X = f(Parents(X))$ per qualche funzione f

P.e., Funzioni Booleane
$NorthAmerican \Leftrightarrow Canadian \vee US \vee Mexican$

P.e., relazioni numeriche tra variabili continue

$$\frac{\partial Level}{\partial t} = \text{inflow} + \text{precipitation} - \text{outflow} - \text{evaporation}$$

Distribuzioni condizionali compatte

Le distribuzioni Noisy-OR modellano cause multiple non interagenti

1) Genitori $U_1 \ldots U_k$ includono tutte le cause (si può aggiungere nodo leak)

2) Probalità q_i di "inibizione" indipendente per ogni causa presa da sola

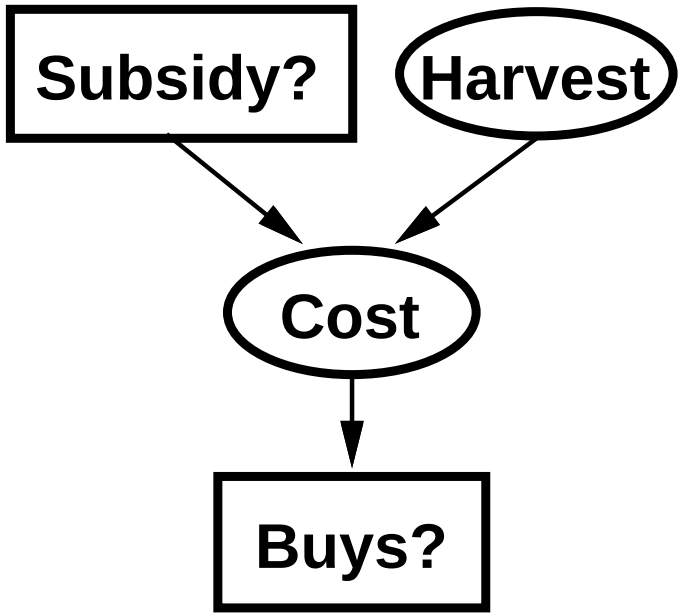
$\Rightarrow P(X|U_1 \ldots U_j, \neg U_{j+1} \ldots \neg U_k) = 1 - \prod_{i=1}^{j} q_i$

$q_1 = 0.6$	$q_2 = 0.2$	$q_3 = 0.1$		
$Cold$	Flu	$Malaria$	$P(Fever)$	$P(\neg Fever)$
F	F	F	**0.0**	1.0
F	F	T	0.9	**0.1**
F	T	F	0.8	**0.2**
F	T	T	0.98	$0.02 = 0.2 \times 0.1$
T	F	F	0.4	**0.6**
T	F	T	0.94	$0.06 = 0.6 \times 0.1$
T	T	F	0.88	$0.12 = 0.6 \times 0.2$
T	T	T	0.988	$0.012 = 0.6 \times 0.2 \times 0.1$

Numero di parametri **lineare** nel numero di genitori

Reti ibride (var. discrete+continue)

Discrete (*Subsidy?* e *Buys?*); continue (*Harvest* e *Cost*)



Opzione 1: discretizzazione — errori possibilmente grandi, CPT grandi
Opzione 2: famiglie canoniche finitamente parametrizzate
1) Variabili continue, genitori discreti+continui (p.e., *Cost*)
2) Variabili discrete, genitori continui (p.e., *Buys?*)

Variabili figlio continue

Occorre una funzione di densità condizionale per la variabile figlio dati i genitori continui, per ogni possibile assegnamento a genitori discreti

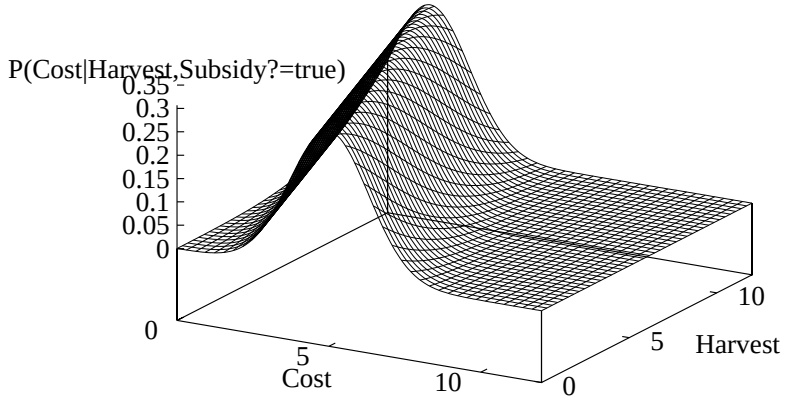
Molto comune è il modello Gaussiano lineare (LG), p.e.,:

$$\begin{aligned} & P(Cost = c | Harvest = h, Subsidy? = true) \\ &= N(a_t h + b_t, \sigma_t)(c) \\ &= \frac{1}{\sigma_t \sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{c - (a_t h + b_t)}{\sigma_t}\right)^2\right) \end{aligned}$$

Indica che $Cost$ varia linearmente con $Harvest$; la varianza è fissata

La variazione lineare non è ragionevole sull'intero intervallo (il costo può assume valori negativi!!) ma va bene se l'intervallo di variabilità **più probabile** di $Harvest$ è stretto

Variabili figlio continue



In una rete con tutte variabili continue modellate con distribuzioni LG

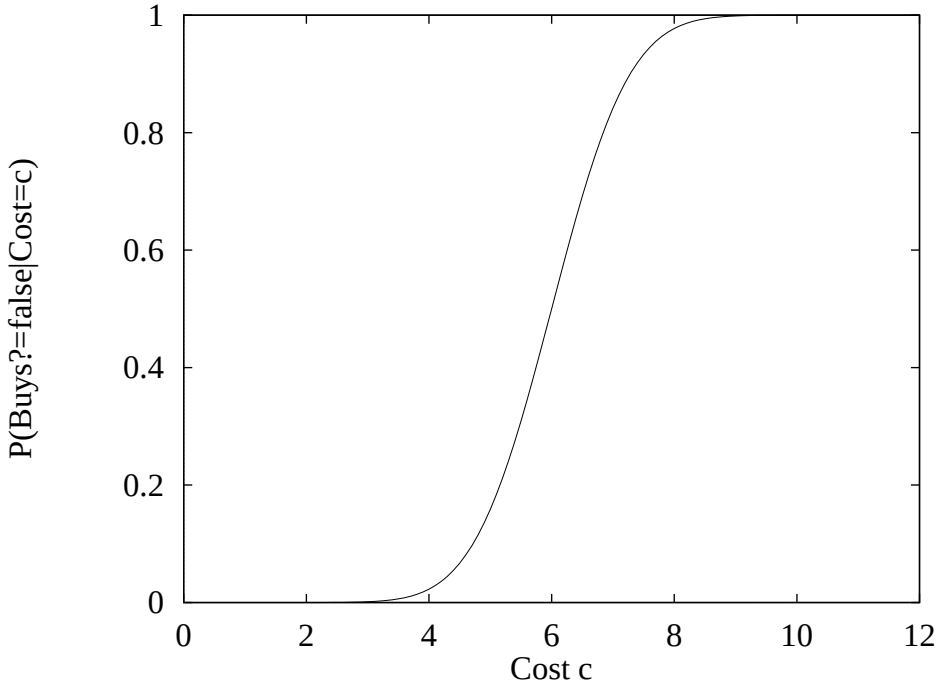
$\Rightarrow$ la distribuzione congiunta completa è una multivariata Gaussiana

Come sopra + variabili discrete $\Rightarrow$ rete condizionale Gaussiana

cioè, una multivariata Gaussiana su tutte le variabili continue per ogni possibile combinazione di valori per le variabili discrete

Variabili discrete con genitori continui

Probabilità di $Buys?$ dato $Cost$ dovrebbe "rappresentare" una soglia "soft":



La distribuzione probit usa l'integrale di una Gaussiana:

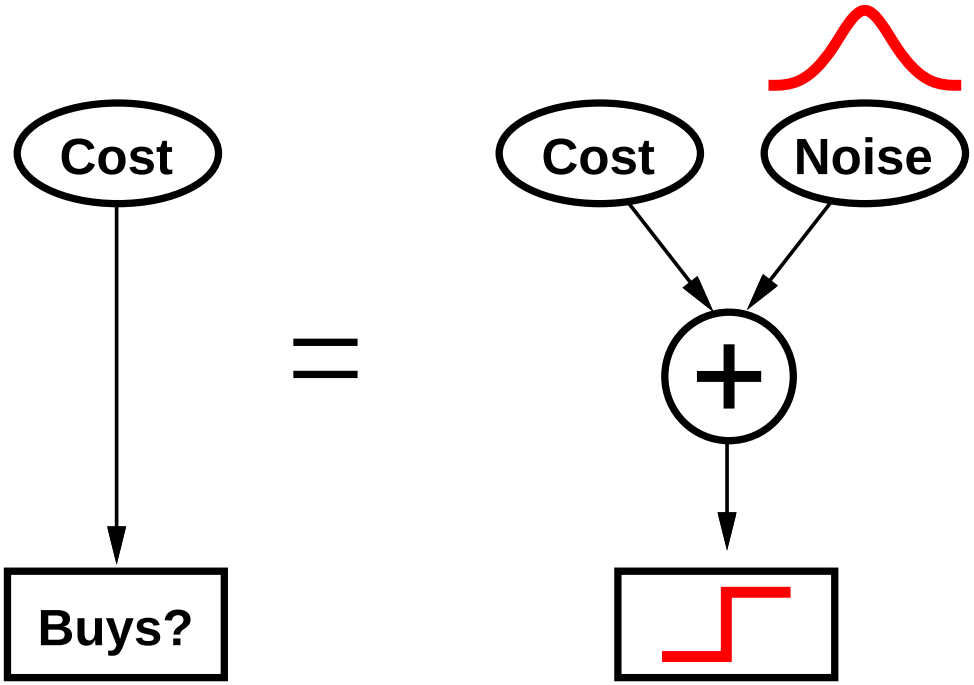
$\Phi(x) = \int_{-\infty}^{x} N(0,1)(x)dx$

$P(Buys? = true \mid Cost = c) = \Phi((-c + \mu)/\sigma)$

Perché usare probit?

1. Ha la forma "giusta"

2. Può essere vista come una funzione soglia la cui localizzazione è soggetta a rumore

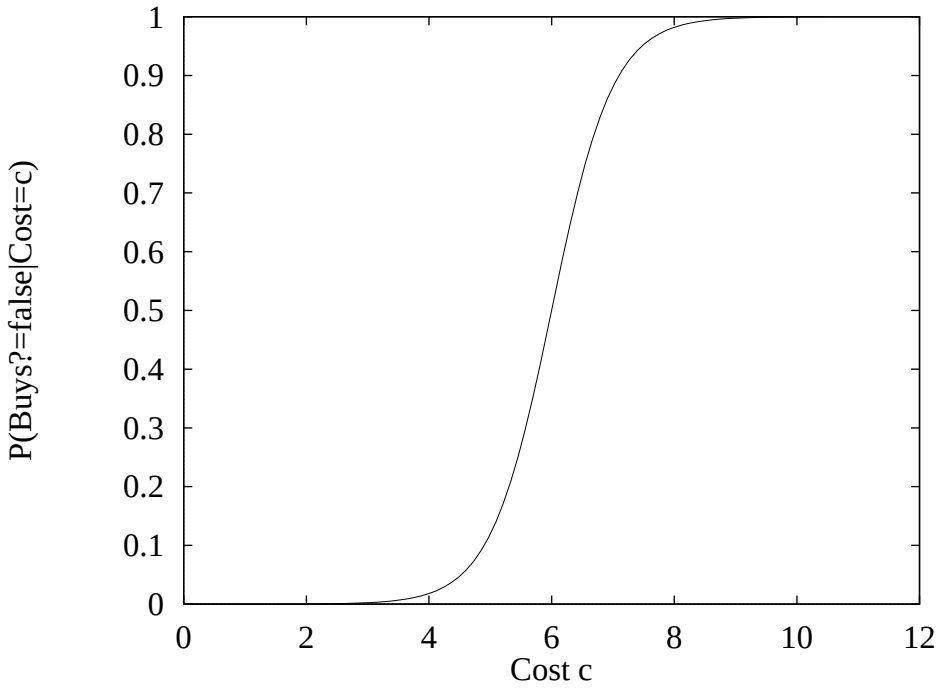


Variabili discrete

Alternativa a probit: la distribuzione sigmoidale (o logit) (usata in reti neurali):

$$P(Buys? = true \mid Cost = c) = \frac{1}{1 + exp(-2\frac{-c+\mu}{\sigma})}$$

La sigmoidale ha forma simile a probit, ma con code più lunghe:



Riassunto

Le Reti Bayesiane forniscono una rappresentazione naturale per l'indipendenza condizionale (indotta causalmente)

Topologia + CPT = rappresentazione compatta della distribuzione congiunta

Solitamente facili da costruire per (non)esperti

Distribuzioni canoniche (p.e., noisy-OR) = rappresentazione compatta di CPT

Variabili continue $\Rightarrow$ distribuzioni parametrizzate (p.e., Gaussiana lineare)