

Intelligenza Artificiale II

Ragionamento probabilistico

Rappresentazione

Marco Piastra

Parte 1

Mondi possibili, sottoinsiemi, eventi

Partizioni e variabili aleatorie

Probabilità

Marginalizzazione

Condizionali

Indipendenza, indipendenza condizionale

Modelli grafici

Eventi

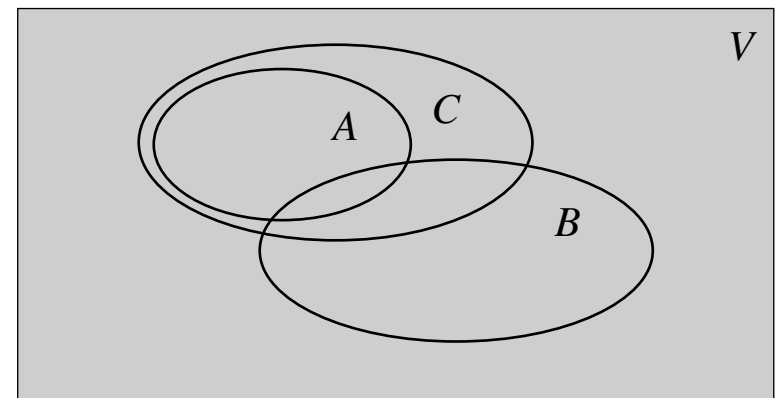
- Logica, formule ed eventi

Si consideri un linguaggio proposizionale L_p

L'insieme dei *mondi possibili* coincide con l'insieme delle interpretazioni V

A ciascun simbolo proposizionale corrisponde un sottoinsieme di V

p.es. ad un simbolo A corrisponde $\{v : v(A) = 1\}$



Quindi, come noto, anche a ciascuna fbf corrisponde un sottoinsieme di V

Un **evento** è un sottoinsieme di *mondi possibili*

Un evento si **verifica** quando il *mondo attuale* appartiene al corrispondente sottoinsieme

Ma l'agente usa le *descrizioni* (fbf) degli eventi, non sa qual'è il mondo attuale

Possibilità

■ Conoscenze certe

L'agente possiede un sistema di conoscenze oggettive, descritto da una teoria

p.es. l'agente sa che $B \vee C$

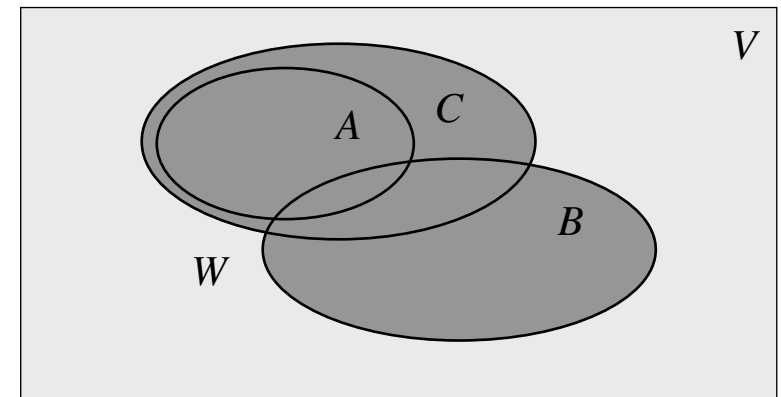
quindi solo i mondi $\{v : v(B \wedge C) = 1\}$ sono *possibili*

Potremmo dire che l'evento $B \wedge C$ si è già *verificato*

Viceversa, il valore di verità di A non è noto (all'agente), infatti:

$$B \wedge C \not\models A$$

$$B \wedge C \not\models \neg A$$



Un **evento** è *possibile* quando non è noto il valore di verità.

L'evento più generico è l'insieme dei mondi possibili

$$W \equiv \{ B \wedge C \}$$

Probabilità*

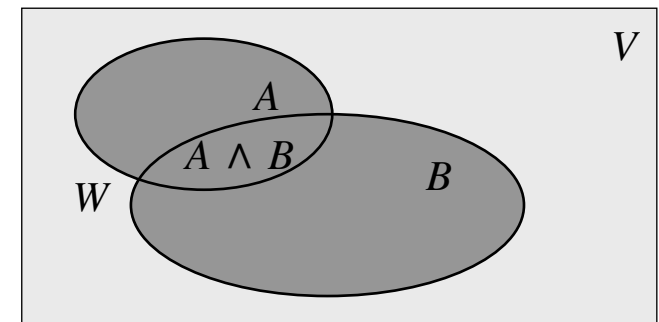
- Una misura dei sottoinsiemi di W

$P(.)$ è una *funzione* che assegna un numero reale a ciascun sottoinsieme di W

- Assiomi (Kolmogorov)

- 1) Per qualsiasi sottoinsieme E , $P(E) \geq 0$
- 2) $P(W) = 1$
- 3) Per qualsiasi sequenza E_i di eventi disgiunti,

$$P(E_1 \vee E_2 \vee \dots \vee E_n) = \sum_i P(E_i)$$



Esempio: $A \wedge \neg B$, $A \wedge B$, $\neg A \wedge B$ sono tre eventi disgiunti (vedi figura)

- $P(A) = P(A \wedge \neg B) + P(A \wedge B)$
- $P(B) = P(\neg A \wedge B) + P(A \wedge B)$
- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

Partizioni, variabile aleatoria*

■ Partizione

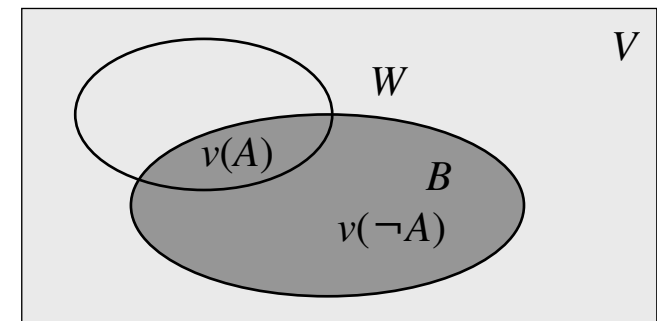
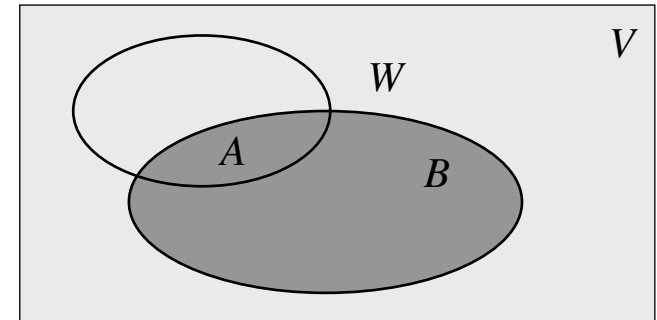
Ciascun simbolo proposizionale A suddivide V in due sottoinsiemi disgiunti, A e $\neg A$

(Quindi $P(\neg A) + P(A) = P(W) = 1$, da cui $P(\neg A) = 1 - P(A)$)

In altri termini:

i possibili valori $v(A) = 1$ e $v(A) = 0$

definiscono una *partizione* di V in base ad A



■ Variabile aleatoria

Si consideri una variabile X che ha $\{x_1, x_2, \dots, x_n\}$ come *dominio*

In ciascun mondo possibile X assume un determinato valore x_i

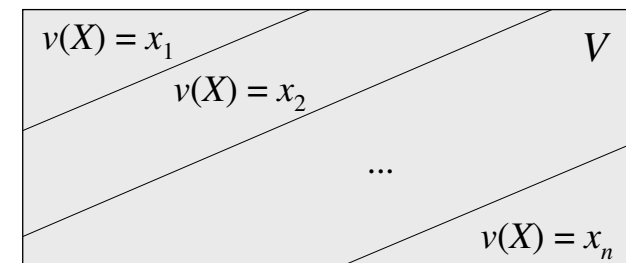
I possibili valori $v(X) = x_1, v(X) = x_2, \dots, v(X) = x_n$ definiscono una *partizione* di V in base ad X

- X è una *variabile aleatoria*
- Ciascun $v(X) = x_i$ è un evento (un sottoinsieme di V)

Anche A (a valori binari) è una variabile aleatoria

Le v.a. binarie o *binomiali* sono anche dette *bernoulliane*

Le v.a. a più valori sono dette *multinomiali*



Partizioni, variabile aleatoria*

■ Partizione

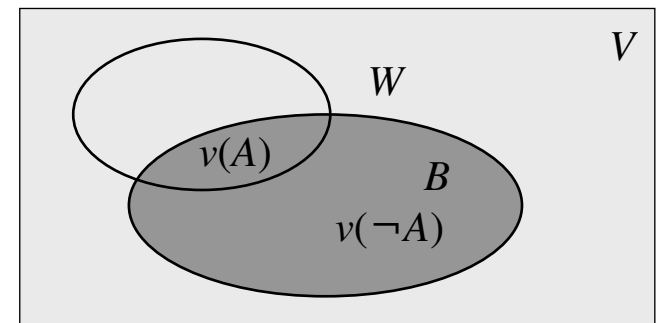
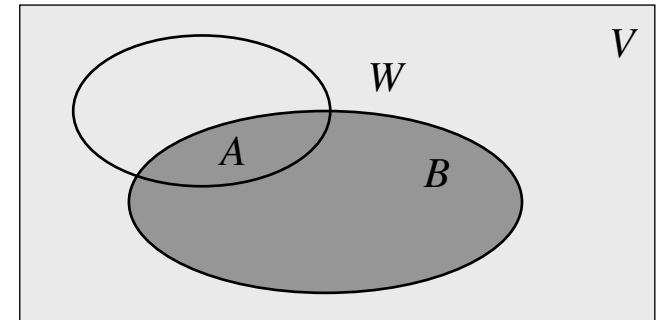
Ciascun simbolo proposizionale A suddivide V in due sottoinsiemi disgiunti, A e $\neg A$

(Quindi $P(\neg A) + P(A) = P(W) = 1$, da cui $P(\neg A) = 1 - P(A)$)

In altri termini:

i possibili valori $v(A) = 1$ e $v(A) = 0$

definiscono una *partizione* di V in base ad A



■ Variabile aleatoria

Si consideri una variabile X che ha $\{x_1, x_2, \dots, x_n\}$ come *dominio*

In ciascun mondo possibile X assume un determinato valore x_i

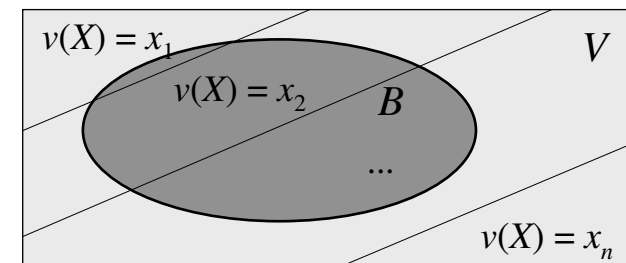
I possibili valori $v(X) = x_1, v(X) = x_2, \dots, v(X) = x_n$ definiscono una *partizione* di V in base ad X

- X è una *variabile aleatoria*
- Ciascun $v(X) = x_i$ è un evento (un sottoinsieme di V)

Anche A (a valori binari) è una variabile aleatoria

Le v.a. binarie o *binomiali* sono anche dette *bernoulliane*

Le v.a. a più valori sono dette *multinomiali*



Variabili aleatorie, distribuzione congiunta*

Essendo $X=x_i$ e $X=x_j$ eventi disgiunti: $P(X=x_i \vee X=x_j) = P(X=x_i) + P(X=x_j)$ se $i \neq j$

■ Variabili aleatorie multiple

Solitamente, in una rappresentazione probabilistica convivono più variabili aleatorie

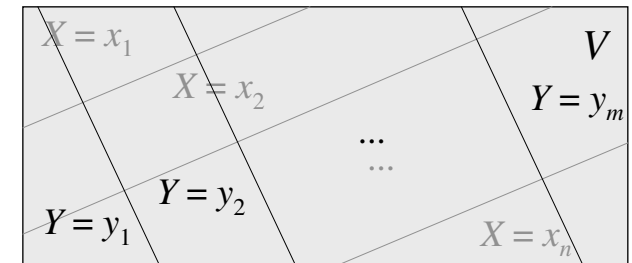
Esempi:

X_i occorrenza in un'email di una parola i

Y classificazione della stessa email come spam

Ciascuna combinazione di valori delle v.a. è un *evento*

Un'insieme di v.a. definisce una partizione di V



■ Distribuzione di probabilità congiunta (*joint probability distribution*)

Per un determinato insieme di variabili aleatorie, p.es. X, Y, Z

E' una funzione $P(X=x_i \wedge Y=y_j \wedge Z=z_k)$ che associa un numero reale a ciascuna combinazione di valori $\langle x_i, y_j, z_k \rangle$

Si indica anche con $P(X=x_i, Y=y_j, Z=z_k)$ oppure $P(X, Y, Z)$

Dato che X, Y e Z definiscono una partizione di V :
$$\sum_i \sum_j \sum_k P(X=x_i, Y=y_j, Z=z_k) = 1$$

Marginalizzazione*

L'eliminazione di una variabile aleatoria da una probabilità congiunta

Data una probabilità congiunta

$$P(X=x_i, Y=y_j, Z=z_k)$$

La *probabilità marginale* $P(X=x_i, Y=y_j)$ si ottiene per sommatoria:

$$P(X=x_i, Y=y_j) = \sum_k P(X=x_i, Y=y_j, Z=z_k)$$

Data una probabilità congiunta su una partizione,
si può sempre ottenere una probabilità congiunta su una partizione contenuta nella prima

Esempio: distribuzione congiunta

(*Vedi anche DutchBook.xls)

La conoscenza della distribuzione di probabilità congiunta permette di stabilire la probabilità di qualsiasi combinazione logica di eventi

Esempi:

$$P(A \vee C) = \sum_B P(A \vee C, B) = 0.55$$

(0.55 * x) dovrebbe essere la somma che siete disposti a scommettere per una vincita x

$$P(\neg A \wedge \neg B) = \sum_C P(\neg A \wedge \neg B, C) = 0.22$$

A	B	C	$P(A, B, C)$
0	0	0	0.10
0	0	1	0.12
0	1	0	0.35
0	1	1	0.08
1	0	0	0.01
1	0	1	0.02
1	1	0	0.23
1	1	1	0.09

Interludio: possibilità e probabilità

- Come si rappresentano i mondi *logicamente* impossibili?

Esempio: ‘testa o croce’ vuol dire che escludiamo ‘testa e croce’, ‘né testa né croce’

a) I mondi *logicamente* impossibili sono esclusi da W

b) I mondi *logicamente* impossibili sono in W ma hanno misura nulla ($P = 0$)

- **Preferenza** (in questo corso)

Alla De Finetti (1930)

I mondi impossibili sono esclusi da W

Come in una struttura modale KD45:

La probabilità è definita sull’insieme dei mondi accessibili

L’agente distingue tra fatti e certezze probabilistiche (= ‘in tutti i mondi possibili’)

In tutti i mondi possibili sono vere le formule logiche che descrivono i fatti noti

In ‘testa o croce’, l’insieme delle conoscenze oggettive è $W \equiv \{\neg(T \wedge C), T \vee C\}$

Probabilità condizionale

■ Definizione

$$P(A|B) = \frac{P(A, B)}{P(B)}$$

■ Significato

E' una forma di *inferenza*: si passa da un'insieme di mondi possibili ad un altro

Quindi, da una misura di probabilità ad un'altra

Si assuma un agente consideri W come insieme di mondi possibili

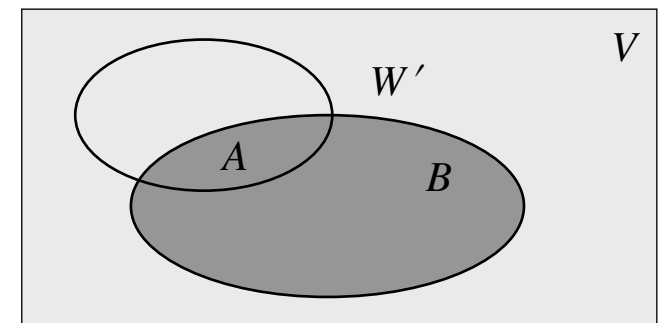
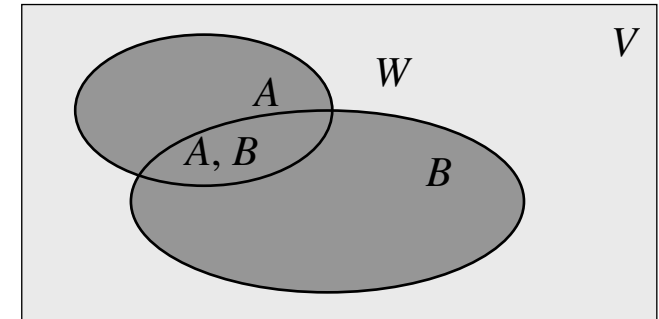
$P(A)$ è la probabilità che A si verifichi

Si supponga che l'agente venga a sapere che l'evento B si è verificato

L'evento complementare $\neg B$ è quindi *impossibile*

$W' \equiv B$ è il nuovo insieme dei mondi possibili

$P(A|B)$ è la nuova probabilità
che l'evento A si verifichi



Esempio: probabilità condizionale

La conoscenza della distribuzione di probabilità congiunta permette di stabilire qualsiasi probabilità condizionale

Esempio:

$$P(A \vee C | B) = \frac{P(A \vee C, B)}{P(B)} = \frac{0.40}{0.75} = 0.53$$

$P(A \vee B)$ era 0.55: la conoscenza di B diminuisce la somma che siete disposti a scommettere

A	B	C	$P(A, B, C)$
0	0	0	0.10
0	0	1	0.12
0	1	0	0.35
0	1	1	0.08
1	0	0	0.01
1	0	1	0.02
1	1	0	0.23
1	1	1	0.09



Teorema di Bayes (T. Bayes, 1764)

▪ Definizione

Una relazione tra probabilità condizionali e marginali

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

Nelle applicazioni pratiche, $P(B|A)$ viene anche detta verosimiglianza (*likelihood*) $L(A|B)$

$$P(A|B) \propto L(A|B) P(A)$$

Corollario della definizione di probabilità condizionale (*chain rule*)

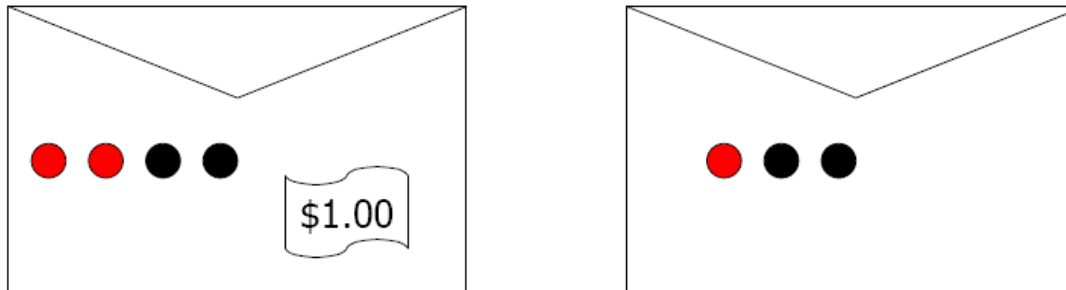
$$P(A, B) = P(B|A) P(A)$$

Per la definizione di marginalizzazione: $P(B) = \sum_A P(A, B) = \sum_A P(B|A) P(A)$

Da cui (formulazione alternativa del teorema di Bayes):

$$P(A|B) = \frac{P(B|A) P(A)}{\sum_A P(B|A) P(A)}$$

Esercizio: informazioni e scommesse



- Due buste, una viene estratta

Una busta contiene due gettoni rossi e due neri, vale 1 euro

Una busta contiene un gettone rosso e due neri, non vale nulla

Prima di scommettere, potete estrarre un gettone

a) Il gettone è nero. Quanto scommettete?

b) Il gettone è rosso. Quanto scommettete?

Obiettivo: mostrare che il teorema di Bayes semplifica i calcoli

Indipendenza, indipendenza condizionale

■ Indipendenza

Due eventi sono indipendenti se la probabilità congiunta è uguale al prodotto delle probabilità marginali

$$\langle A \perp B \rangle \Rightarrow P(A, B) = P(A) P(B)$$

■ Indipendenza condizionale

Due eventi sono condizionalmente indipendenti (dato un terzo evento) se la probabilità condizionale congiunta è uguale al prodotto delle probabilità condizionali marginali

$$\langle A \perp B | C \rangle \Rightarrow P(A, B | C) = P(A | C) P(B | C)$$

$$\Rightarrow P(A | B, C) = \frac{P(A, B | C)}{P(B | C)} = \frac{P(A | C) P(B | C)}{P(B | C)} = P(A | C)$$

Questa è la proprietà più rilevante

Modelli grafici (anche *Bayesian Networks*)

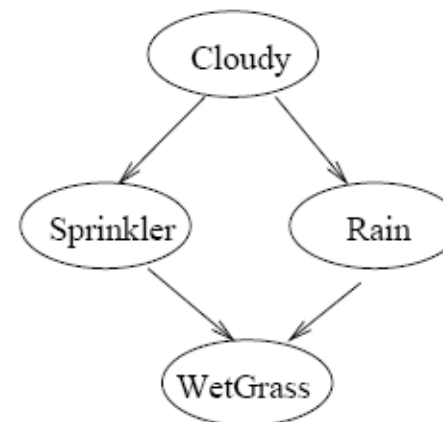
Struttura + numeri, invece di soli numeri

- Un modo per rappresentare una distribuzione di probabilità congiunta

I nodi sono variabili aleatorie

Gli archi (orientati) rappresentano *dipendenza*

C	P(S=F)	P(S=T)
F	0.5	0.5
T	0.9	0.1



C	P(R=F)	P(R=T)
F	0.8	0.2
T	0.2	0.8

Notare che la specifica di una distribuzione congiunta di quattro v.a. richiederebbe $2^4 = 16$ valori

In figura i valori sono solo 9

S	R	P(W=F)	P(W=T)
F	F	1.0	0.0
T	F	0.1	0.9
F	T	0.1	0.9
T	T	0.01	0.99

Da un modello grafico alla probabilità congiunta

■ Distribuzione congiunta

Può essere espressa come prodotto di probabilità condizionali

(estensione della *chain rule*)

Esempio:

$$P(C, S, R, W) = P(C)P(S|C)P(R|S,C)P(W|R,S,C)$$

In un modello grafico, la distribuzione congiunta è un prodotto delle probabilità condizionali dei nodi

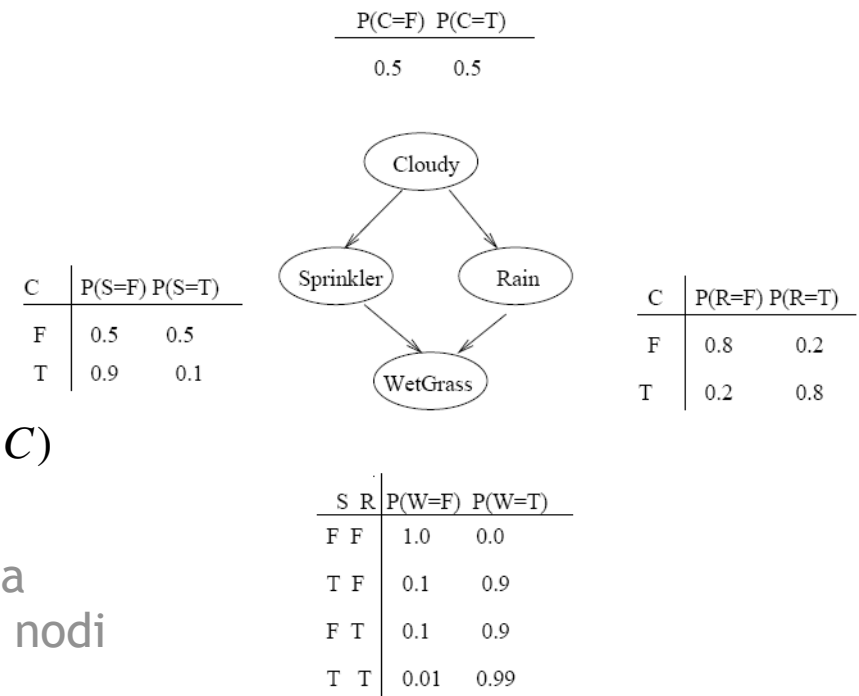
$$P(X_1, X_2, \dots, X_n) = \prod_i P(X_i | \text{parents}(X_i))$$

Dove $\text{parents}(X_i)$ sono i nodi afferenti (diretti) del grafo orientato

Nell'esempio:

$$P(C, S, R, W) = P(C)P(S|C)P(R|C)P(W|R,S)$$

Assunzioni implicite: $\langle R \perp S | C \rangle$, $\langle W \perp C | R, S \rangle$



Modello grafico e indipendenze condizionali

■ *D-separation (Dependency-separation)*

Come si 'legge' l'indipendenza condizionale in un modello grafico

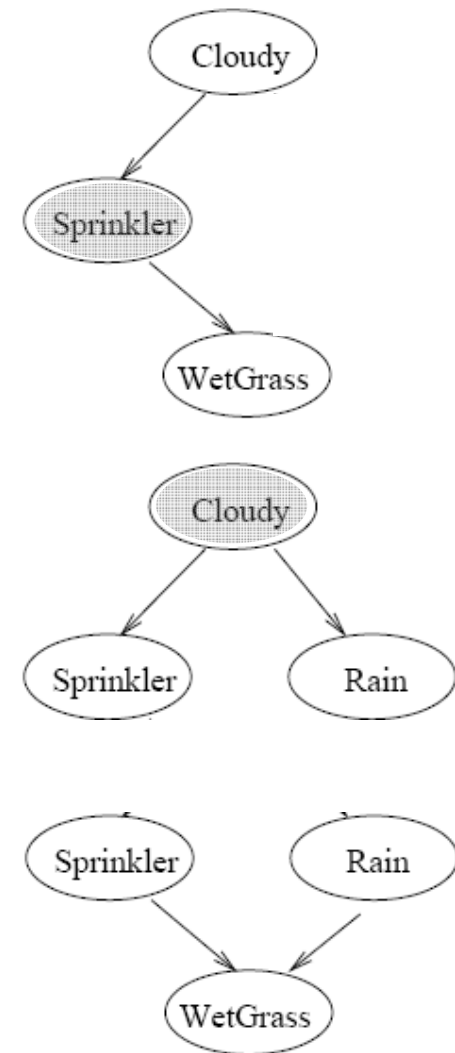
In un modello grafico

Due nodi X e Y sono condizionalmente indipendenti dato un insieme di nodi $\{Z_k\}$ se tutti i percorsi tra X e Y sono bloccati

Nel determinare i possibili percorsi tra due nodi, si ignora il verso degli archi

Un percorso tra X e Y è bloccato se:

- 1) Il percorso contiene una sequenza $X \rightarrow Z_i \rightarrow Y$ oppure una diramazione (*fork*) $X \leftarrow Z_i \rightarrow Y$ ($Z_i \in \{Z_k\}$)
- 2) Il percorso contiene una confluenza (*join*) $X \rightarrow N \leftarrow Y$ in cui N e tutti i discendenti di N non appartengono a $\{Z_k\}$



Explaining Away

Ulteriori osservazioni sulla condizione 2) della *D-separation*

Modello grafico con un *join*

Probabilità congiunta, dal grafo:

$$P(X, Y, Z) = P(X)P(Y)P(Z|X, Y)$$

Probabilità marginale rispetto a X e Y (valore di Z incognito):

$$P(X, Y) = P(X)P(Y) \sum_Z P(Z|X, Y) = P(X)P(Y)$$

Quindi X e Y sono marginalmente indipendenti

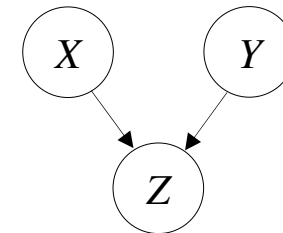
Ma se il valore di Z è noto, allora X e Y sono dependenti:

$$P(X, Y | Z=v) = \frac{P(X, Y, Z=v)}{P(Z=v)} = \frac{P(X)P(Y)P(Z=v|X, Y)}{\sum_{X, Y} P(X)P(Y)P(Z=v|X, Y)}$$

Non è un paradosso.

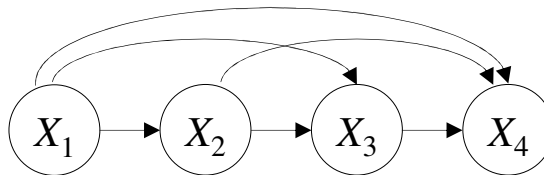
Esempio:

X e Y sono due lanci della stessa moneta, $Z=1$ se il risultato è lo stesso, $Z=0$ altrimenti.



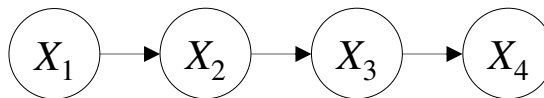
Esempi di modelli grafici

▪ Dipendenza completa



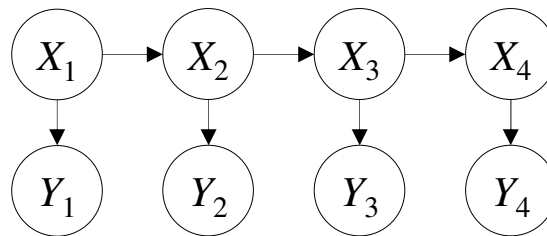
$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2 | X_1)P(X_3 | X_1, X_2)P(X_4 | X_1, X_2, X_3)$$

▪ Modello di Markov



$$P(X_1, X_2, X_3, X_4) = P(X_1)P(X_2 | X_1)P(X_3 | X_2)P(X_4 | X_3) = P(X_1) \prod_{i=2}^n P(X_i | X_{i-1})$$

▪ Modello 'Hidden Markov'



In genere, i nodi X_i sono *hidden*, nel senso di *non-osservabili*

$$\begin{aligned} P(X_1, X_2, X_3, X_4, Y_1, Y_2, Y_3, Y_4) &= P(X_1)P(Y_1 | X_1)P(X_2 | X_1)P(Y_2 | X_2)P(X_3 | X_2)P(Y_3 | X_3)P(X_4 | X_3)P(Y_4 | X_4) \\ &= P(X_1)P(Y_1 | X_1) \prod_{i=2}^n P(X_i | X_{i-1})P(Y_i | X_i) \end{aligned}$$