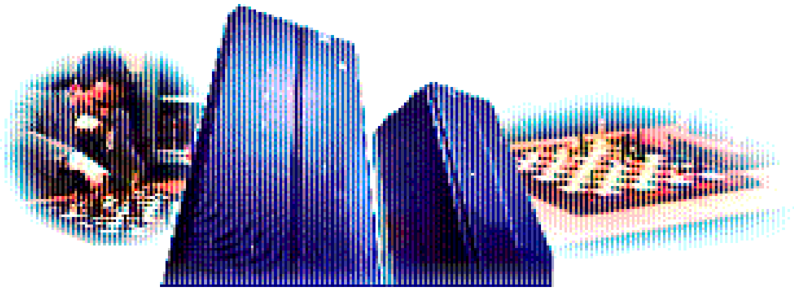
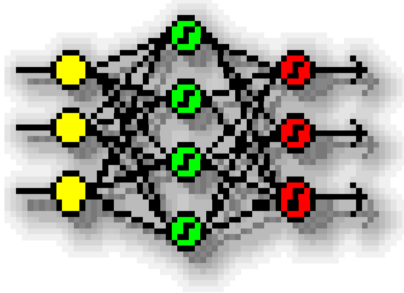




Sistemi Intelligenti
Introduzione al calcolo delle probabilità - I

Alberto Borghese

Università degli Studi di Milano
Laboratory of Applied Intelligent Systems (AIS-Lab)
Dipartimento di Informatica
borghese@di.unimi.it

Overview

Probabilità semplice e condizionata

Esempio

Teorema di Bayes

Incertezza

- Le azioni “intelligenti” vengono fatte verso un ambiente che presenta una certa dose di **incertezza**.

E.g. Dobbiamo andare a Malpensa. Quanto tempo prima dobbiamo partire?
Dalla nostra esperienza deriviamo che 60 minuti sono sufficienti se.....

Rimane un po’ di incertezza. Se partiamo 120 minuti prima ci teniamo un margine, ma passeremo facilmente tanto tempo in aereoporto senza fare nulla.

Quando prendiamo una decisione, teniamo conto in modo più o meno esplicito di questi elementi. Questi elementi hanno a che fare con a statistica.

Probabilità (visione frequentista)

$$P(A=a_i) = \lim_{N->\infty} \frac{n_{A=a_i}}{N} = \lim_{N->\infty} \frac{n_i}{N}$$

Per il teorema del limite centrale la frequenza di un evento su infinite realizzazioni è uguale alla sua probabilità.

Supponiamo A = {a_1, a_2}

La probabilità che si verifichi uno tra tutti i casi possibili è sempre 1.

Ovverosia la somma delle probabilità di tutti gli eventi (se mutuamente esclusivi) somma 1.

$$P(A=a_1) \cup P(A=a_2) = \lim_{N->\infty} \frac{n_{A=a_1}}{N} + \lim_{N->\infty} \frac{n_{A=a_2}}{N} = \lim_{N->\infty} \frac{n_{A=a_1} + n_{A=a_2}}{N} = 1$$

$$P(A) = P(A=a_1) + P(A=a_2) = 1$$

Altri aspetti della probabilità

Problema della visione frequentista: **omogeneità del campione** (classe di riferimento). Come posso effettuare la media di eventi in modo “sicuro”?

- **Visione oggettivista.** Tendenza di un fenomeno ad accadere. Se lanciamo una moneta in aria, possiamo affermare che avremo 50% di probabilità che esca testa e 50% che esca croce. Ci aspettiamo che questa affermazione venga supportata quando effettuiamo infiniti esperimenti.
- **Visione soggettivista.** La probabilità viene espressa come credenza del soggetto. “Secondo me la probabilità di avere una carie è del 10%”. Non dipendono da un ragionamento fisico e rappresentano una probabilità a-priori. Deve potere essere corretta quando arrivano evidenze sperimentali.

Esempio di incertezza

Mal di denti => Carie   Quando possiamo essere sicuri che questa proposizione (evento) sia vera?
(*notice that proposition is used in logic, it is an event in AI terminology*).

In realtà non è sempre vera: il mal di denti può avere diverse **cause**

Mal di denti => Carie OR Problemi gengive OR ascessi OR .....

Vale il viceversa?   Carie => Mal di denti

Quali sono i problemi con l'approccio puramente logico?

Laziness (svogliatezza). Non si riescono ad elencare tutte le situazioni associate al mal di denti

Ignoranza teorica. Non abbiamo una conoscenza che spieghi tutto nel dominio di interesse.

Ignoranza pratica. Anche se avessimo una conoscenza completa, non riusciamo a conoscere le condizioni esatte in cui si verifica l'evento (paziente).

Possiamo ottenere un **grado di credenza (belief)** nell'affermazione. Questa potrà rivelarsi vera o falsa con una certa probabilità.

La probabilità è basata sulla conoscenza (a-priori) non sull'evento che si è già verificato!!

Formulazione analitica

Gli eventi sono mutuamente esclusivi e costituiscono lo spazio delle probabilità: Ω, dove ciascun evento viene indicato con ω.

Questo porta a definire gli assimi di base della teoria della probabilità:

Quindi:

1. $0 \leq P(\omega) \leq 1$ per ogni ω

$$\sum_{\omega \in \Omega} P(\omega) = 1$$

2. P(a) = True $\longleftrightarrow$ P(!a) = False
3. P(a OR b) = P(a) + P(b) – P(a AND b) =

   ½ + ½ - ¼ = 3/4

   Principio di inclusione-esclusione

P(a) P(b)

a	b	a OR b
0	0	0
0	1	1
1	0	1
1	1	1

a AND b

P(a AND b) = P(a) P (b) se sono indipendenti

½ ½ = 1/4

Teorema di de Finetti (1931)

- Se un Agente 1 esprime un insieme di credenze (probabilità a-priori) inconsistenti (che violano gli assiomi della teoria delle probabilità), allora ci sarà una combinazione di credenze di un agente 2 che **garantisce** che l'agente 1 perderà.

- Il comportamento di un agente è più efficace quando tutta la conoscenza statistica viene sfruttata in modo congruo (supponendo la conoscenza statistica un-biased).

Questo teorema è stato tradotto in inglese solo nel 1993!

Probabilità a-priori e a-posteriori

Qual'è la probabilità che la proposizione: "E' uscito 11" tirando due dadi si avveri?

P(N=11) = P((5,6) + P(6,5) = 1/6*1/6 + 1/6*1/6 = 1/18

Probabilità incondizionata o a-priori. Non richiede o dipende da altre informazioni.

Supponiamo ora che il primo dado abbia mostrato 5. Abbiamo un'informazione. Perchè N = 11, occorre che il secondo dato mostri 6.

$P(N=11 \mid Dado_1 = 5) = 1/6 \; > \; P(N=11)$ Abbiamo un'incertezza minore.

Probabilità condizionata o a-posteriori.

Un agente cerca di raccogliere più informazioni possibili per diradare l'incertezza e formulare quindi dei problemi descrivibili con probabilità condizionate.

Probabilità condizionata e congiunta

P(a | b) = P(a AND b) / P(b)   P(a AND b) è probabilità congiunta

- P(N = 11 | Dado1 = 5) = P(N=11 AND Dado1 = 5) / P(Dado1 = 5) =
  (1/36) / (1/6) = 1/6

b = Dado1 = 5, restringe le possibili configurazioni. Ne scarta 5/6.

Si può vedere la probabilità condizionata come una funzione a = f(b)

Possiamo riscrivere la probabilità condizionata come:

P(a AND b) = P(a | b) P(b)

Nel caso dei dadi:

P(N=11 AND Dado1 = 5) = P(N = 11 | Dado1 = 5) * P(Dado1 = 5) =

(1/6) * (1/6) = 1/36.

Inferenza statistica

- Calcolo della probabilità a-posteriori di un evento, a partire dall'informazione collezionata sperimentalmente.

- Consideriamo tre variabili binarie: Mal di denti, Carie, Cavità in dente, e le probabilità congiunte:

	mal di denti		!mal di denti	
	cavità	!cavità	cavità	!cavità
carie	0,108	0,012	0,072	0,008
!carie	0,016	0,064	0,144	0,576

$$\sum P(a_i, b_j, c_k) = 1$$

La nostra "funzione" misura il mal di denti e se c'è una cavità (effetto) in dipendenza o meno della presenza di carie (la causa)

Esempi di inferenza statistica

	mal di denti		!mal di denti	
	cavità	!cavità	cavità	!cavità
carie	0,108	0,012	0,072	0,008
!carie	0,016	0,064	0,144	0,576

P(carie OR mal di denti) = 0,108 + 0,012 + 0,072 + 0,008 + 0,108 + 0,012 + 0,016 + 0,064 - 0,108 – 0,012= 0,28

	mal di denti		!mal di denti	
	cavità	!cavità	cavità	!cavità
carie	0,108	0,012	0,072	0,008
!carie	0,016	0,064	0,144	0,576

P(carie) = 0,108 + 0,012 + 0,072 + 0,008 = 0,2

$$P(Y) = \sum_{z \in Z} P(Y, z)$$

Marginalizzazione rispetto a “carie” = Y (summing out): tutte le variabili diverse da “carie”, collassano nella sommatoria.

Condizionamento statistico

	mal di denti		!mal di denti	
	cavità	!cavità	cavità	!cavità
carie	0,108	0,012	0,072	0,008
!carie	0,016	0,064	0,144	0,576

Probabilità a-priori:

P(mal di denti) = 0,108 + 0,012 + 0,016 + 0,064 = 0,2

P(!mal di denti) = 0,8

Probabilità condizionate: P(a | b) = P(a AND b) / P(b)

	mal di denti		!mal di denti	
	cavità	!cavità	cavità	!cavità
carie	0,54	0,06	0,09	0,01
!carie	0,08	0,32	0,18	0,72

Probabilità condizionate P(carie| mal di denti)

Condizionamento statistico

	mal di denti		!mal di denti	
	cavità	!cavità	cavità	!cavità
carie	0,54	0,06	0,09	0,01
!carie	0,08	0,32	0,18	0,72

Probabilità condizionate
P(carie| mal di denti)

P(mal di denti) = 0,108 + 0,012 + 0,016 + 0,064 = 0,2
P(!mal di denti) = 0,8

$$P(carie) = P(Y) = \sum_{z \in Z} P(Y \mid z) P(z)$$ = P(Carie | mal di denti)P(mal di denti) +

P(Carie | !mal di denti)P(!mal di denti) = (0,54+0,06) * 0,2 + (0,09+0,01) * 0,8 = 0,2

Come utilizziamo la stima a-posteriori

Vogliamo determinare la probabilità di un evento partendo dalla conoscenza sperimentale di altre.

	mal di denti		!mal di denti	
	cavità	!cavità	cavità	!cavità
carie	0,108	0,012	0,072	0,008
!carie	0,016	0,064	0,144	0,576

P(carie | mal di denti) = P(carie AND mal di denti) / P(mal di denti)

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

P(!carie | mal di denti) = P(!carie AND mal di denti) / P(mal di denti) = 0.4

P(mal di denti) è la probabilità marginale relativa al mal di denti P(mal di denti) = 0,2. Ha una funzione di normalizzazione. Il rapporto tra P(carie) e P(!carie) non dipende da P(mal di denti).

P(carie | mal di denti) = α P(carie AND mal di denti)

Inferenza statistica nel caso generale

- Consideriamo una funzione con incertezza: P(X | E, Y) dove e sono le variabili osservate e y quelle non osservate.

- La P(X | e) = P(X,e) = $\sum_{y \in Y} \left(P(X, e, y) / P(y) \right)$

Quando si vuole una valutazione comparativa, il termine P(y) non cambia la funzione P(X|e) e può essere omesso:

P(X | e) = P(X,e) = $\alpha \sum_{y \in Y} P(X, e, y)$

Non si riesce a rappresentare graficamente comunque quando il numero di variabili cresce.

Indipendenza statistica

P(X | Y) = P(X) se X non dipende da Y

P(X, Y) = P(X) P(Y) se X ed Y sono indipendenti.

Overview

Probabilità semplice e condizionata

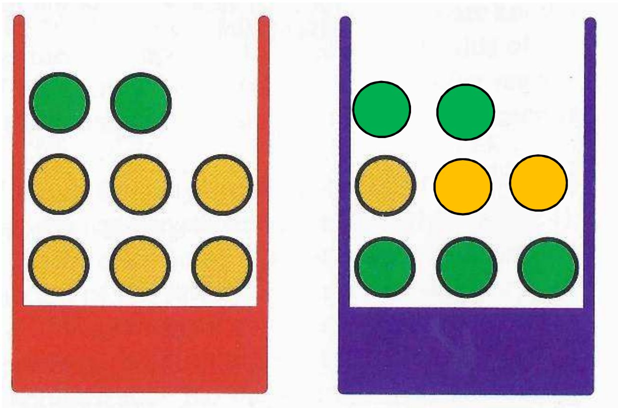
Esempio

Teorema di Bayes

Un altro esempio

Box = {r(ed), b(lue)}

Fruit = {a(pple), o(range)} (apple sono mele verdi)



Supponiamo di prendere un frutto alla volta e di sostituirlo con un frutto dello stesso tipo (sampling with replacement)

Supponiamo di selezionare Box = *r* 40% delle volte e
Box = *b* il 60% delle volte.

P(Box = r) = 0,4 P(Box = b) = 0,6 P(Box) = 1

Formulazione statistica

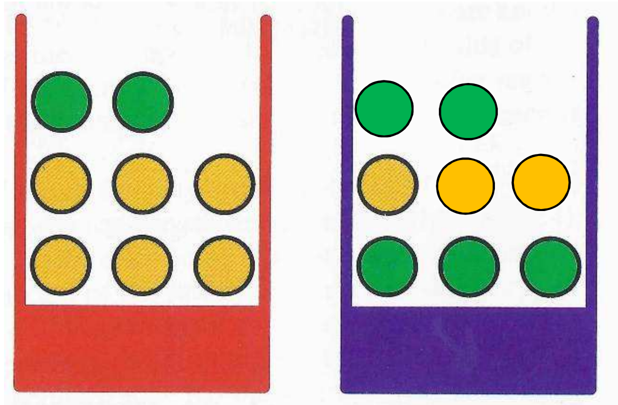
	Red	Blue
Apple		n_{ij}/n_{TOT}
Orange		

y_j

x_i

$X = \{x_i\} = \{\text{blue, red box}\}$ box colour

$Y = \{y_j\} = \{\text{apple, orange}\}$ fruit type



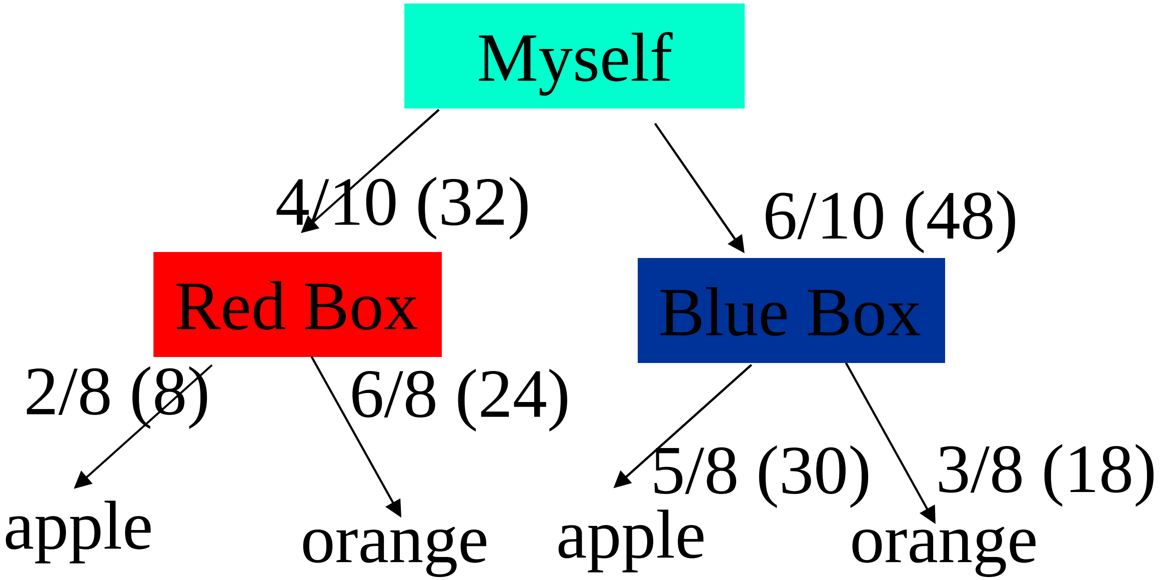
N trials – z_{ij} is the sample observed in each trial (x_{ij}, y_{ij}).

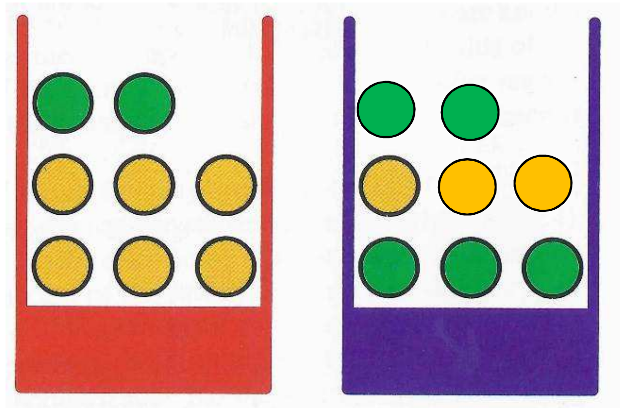
n_{ij} is the number of trials in which x_i AND y_j were observed.

PROBABILITA' CONGIUNTA

$$\sum_i \sum_j \frac{n_{i,j}}{N_{TOT}} = 1$$

Esempio (N=80)





	Red	Blue
Apple	8/80	30/80
Orange	24/80	18/80

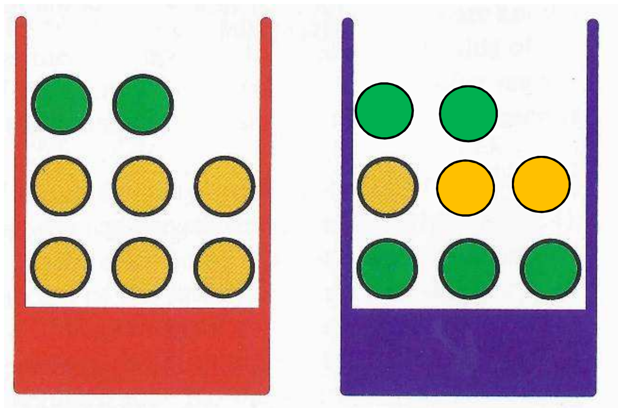
y_j

x_i

$$\sum_{i,j}\frac{n_{i,j}}{N}=1 \quad = \frac{1}{80}\left(8+24+30+18\right)=1$$

Some questions

- What is the probability that the selection procedure will produce an apple?
- What is the probability that the box selected was the blue one if we pick up an orange?



Probabilità congiunta

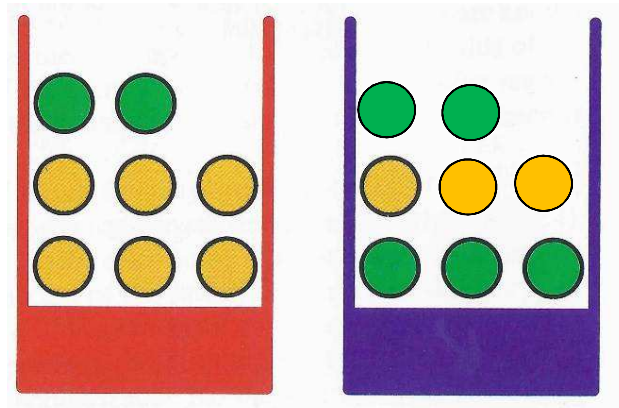
- What is the probability that the **blue** box is chosen **and** an **apple** is selected?

Myself

4/10 (32) — Red Box
6/10 (48) — Blue Box

Red Box: 2/8 (8) apple; 6/8 (24) orange

Blue Box: 5/8 (30) apple; 3/8 (18) orange



	Red	**Blue**
Apple	**4/10*2/8=8/80 – 8**	**6/10*5/8=30/80 – 30**
Orange	4/10*6/8=24/80 - 24	6/10*3/8=18/80 - 18

y_j

x_i

$$\boxed{P(X=x_i;\ Y=y_j)} = \lim_{N \to +\infty} \frac{n_{i,j}}{N}$$

$$\boxed{P(X=b;\ Y=a)} = 30/80 = 3/8$$

Conditional (posterior) probability

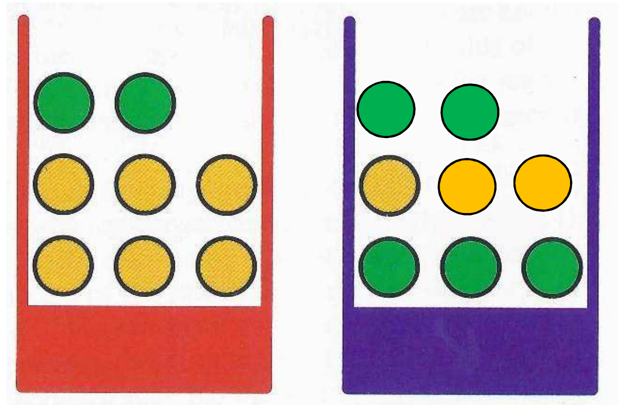
Myself

4/10 (32) | 6/10 (48)

Red Box: 2/8 (8) apple, 6/8 (24) orange

Blue Box: 5/8 (30) apple, 3/8 (18) orange

Apple when the blue box is chosen.



	Red	Blue
Apple	**4/10*2/8=8/80 – 8**	**6/10*5/8=30/80 – 30**
Orange	4/10*6/8=24/80 - 24	6/10*3/8=18/80 - 18

y_j

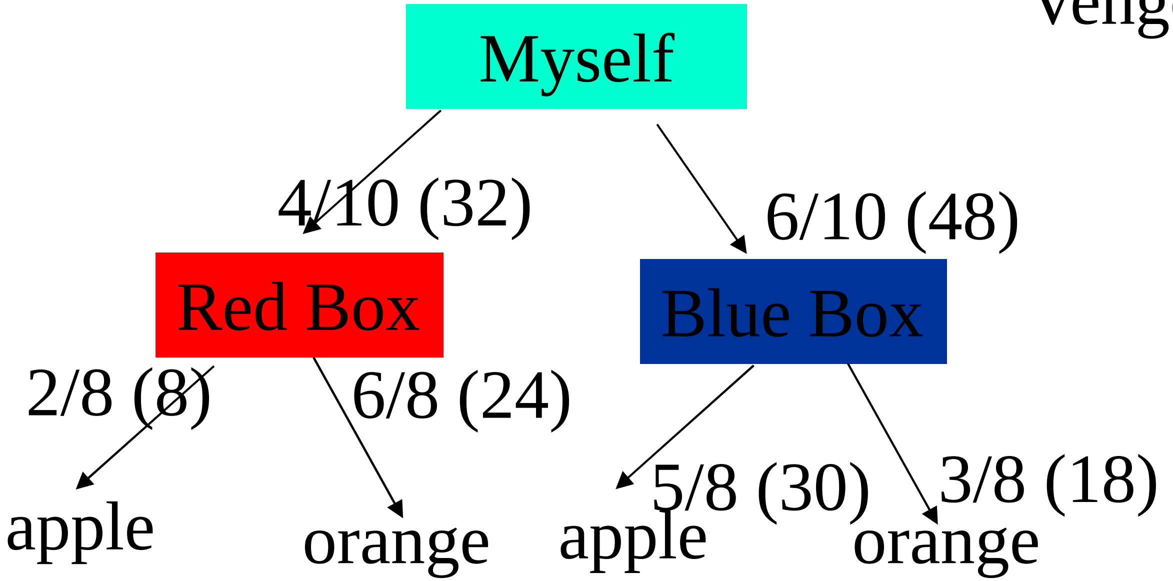
x_i

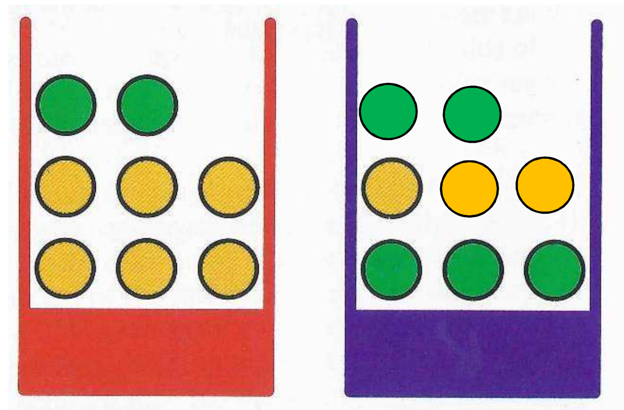
$$\boxed{P(Y=y_j \mid X = x_i)} = \lim_{N \to +\infty} \frac{n_{ij}}{\sum_j n_{ij}}$$

$$\boxed{P(Y=a \mid X = b)} = \frac{n_{a,b}}{n_{a,b} + n_{o,b}} = \frac{30}{30+18} = 5/8$$

Probabilità semplice (marginale)

Vengono eliminate le altre variabili





	Red	Blue
Apple	**4/10*2/8=8/80 – 8**	**6/10*5/8=30/80 – 30**
Orange	4/10*6/8=24/80 - 24	6/10*3/8=18/80 - 18

y_j

x_i

$$\boxed{P(X = r)} = \lim_{N \to +\infty} \frac{\sum_j n_{rj}}{N} = \frac{8}{80} + \frac{24}{80} = \frac{32}{80} = 4/10$$

$$\boxed{P(Y = o)} = \lim_{N \to +\infty} \frac{\sum_i n_{i,o}}{N} = \frac{24}{80} + \frac{18}{80} = \frac{42}{80}$$

$$\boxed{P(Y = a)} = \lim_{N \to +\infty} \frac{\sum_i n_{i,a}}{N} = \frac{8}{80} + \frac{30}{80} = \frac{38}{80}$$

Probabilità marginale

$$\boxed{P(X = r)} = \lim_{N \to +\infty} \frac{\sum_j n_{rj}}{N} = \frac{8}{80} + \frac{24}{80} = \frac{32}{80} = 4/10$$

E' la probabilità ottenuta sommando le probabilità delle altre variabili (mettendo ai margini), in questo caso la variabile Y (frutto)

$$\boxed{P(X = r)} = \lim_{N \to +\infty} \frac{\sum_j n_{r,j}}{N} = \sum_j P(X = r, Y = y_j)$$

$$\boxed{P(Y = o)} = \lim_{N \to +\infty} \frac{\sum_i n_{i,o}}{N} = \sum_j P(X = x_i, Y = o)$$

	Red	Blue
Apple	**4/10*2/8=8/80 – 8**	**6/10*5/8=30/80 – 30**
Orange	4/10*6/8=24/80 - 24	6/10*3/8=18/80 - 18

y_j

x_i

Riassumendo

$$\boxed{P(X=x_i \text{ AND } Y=y_j)} = \lim_{N \to +\infty} \frac{n_{i,j}}{N}$$

Probabilità congiunta

$$\boxed{=P(Y=y_j \mid X = x_i)}\,\boxed{P(X=x_i)}$$

$$\boxed{P(X=x_i)} = \lim_{N \to +\infty} \frac{\sum_j n_{ij}}{N}$$

Probabilità semplice (marginale)

$$\boxed{P(Y=y_j \mid X = x_i)} = \lim_{N \to +\infty} \frac{n_{ij}}{\sum_j n_{ij}}$$

Probabilità condizionata

$$\boxed{=P(Y=y_j \text{ AND } X = x_i)\ /}\,\boxed{P(X=x_i)}$$

Overview

Probabilità semplice e condizionata

Esempio

Teorema di Bayes

Probabilità condizionata

Causa -> Carie trovata dal dentista
Effetto -> mal di denti misurato dal paziente

P(carie | mal di denti)

	mal di denti		!mal di denti	
	cavità	!cavità	cavità	!cavità
carie	0,108	0,012	0,072	0,008
!carie	0,016	0,064	0,144	0,576

Si può pensare come una funzione probabilistica che, dato un valore di mal di denti restituisce un valore di probabilità associato alla variabile carie (molti pazienti vanno dal dentista perchè hanno mal di denti).

Questo è analogo al calcolo di una funzione inversa: dall'effetto risaliamo alla causa. Possiamo dire qualcosa sull'effetto, ma sulla causa?

P(mal di denti | carie) = P(mal di denti, carie) / P(carie) = 0,12 / 0,2 = 0,6 (60% persone che hanno una carie hanno anche mal di denti).

Riusciamo a farlo in pratica? Non abbiamo idea se e quando una carie possa causare un mal di denti. Non sappiamo se abbiamo una carie prima di andare dal dentista! Ma sappiamo che abbiamo mal di denti!!

Teorema di Bayes

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

$$\boxed{P(X|Y) = \frac{P(Y \mid X)P(X)}{P(Y)}}$$

X = causa   Y = effetto

$$\boxed{P(\text{causa}|\text{effetto}) = \frac{P(Effetto|Causa)\,P(Causa)}{P(Effetto)}}$$

We usually do not know the statistics of the cause, but we can measure the effect and , through frequency, build the statistics of the effect or we know it in advance.

A doctor knows P(Symptons|Causa) and wants to determine P(Causa|Symptoms)

Example

Diagnosis (Cause) meningitis

Effect (Symptom) stiff neck

Knowledge a-priori:

1. when meningitis -> stiff neck 70% of the times
2. Meningite incidence 1/50,000
3. Probability of stiff neck 1%

What is the probability that a patient has meningitis if he has stiff neck?

1. P(meningitis | stiff) =
2. P(meningitis) = 0,002
3. P(stiff) = 0,01

From Bayes rule: P(meningitis | stiff) = P(stiff | meningitis)*P(meningitis) / P(stiff) = 0,0014 = 0,14%

Do we need to consider P(stiff)?

We are not interested on the probability of having a stiff neck as we are interested only to what happens since we do find a stiff neck.

For this reason the probability a-priori of the effect can be neglected (in particular in the estimate of the posterior probability) and the conditional probability obtained through Bayes rule and most often considered takes the following shape:

P(Cause | Effect) = α P(Effect | Causa) * P(Causa) α cost

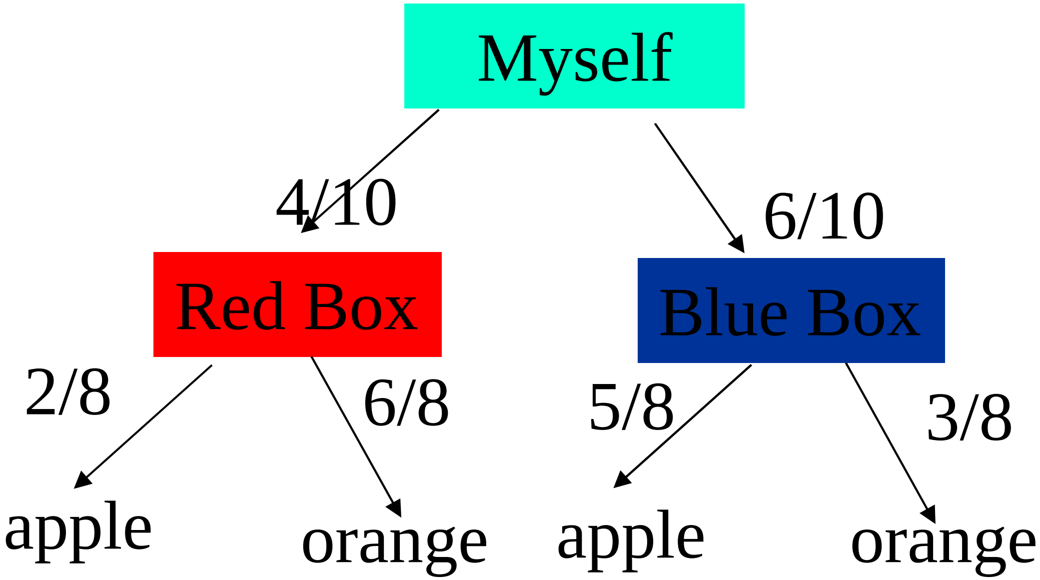
This formulation is interesting as far as we have sufficient knowledge and it is based on empirical evidence. Any doctor knows that 0,14% is the incidence of menigitis with stiff neck. However, what happens if there is an epidemy of meningitis?

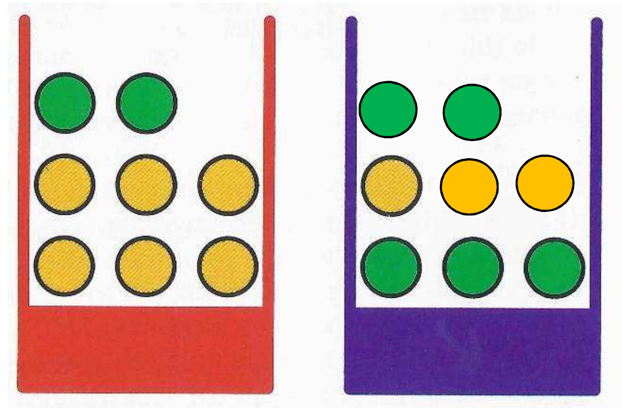
A doctor that uses knowledge not from experience would not change the odds.

A doctor that does use experience would see P(m) raise and therefore would be able to raise P(m | s) in this situation. Notice that P(s | m) does not change.

This makes Bayesian inference one of the most popular tools in intelligent machines and learning.

Il teorema di Bayes nel nostro caso



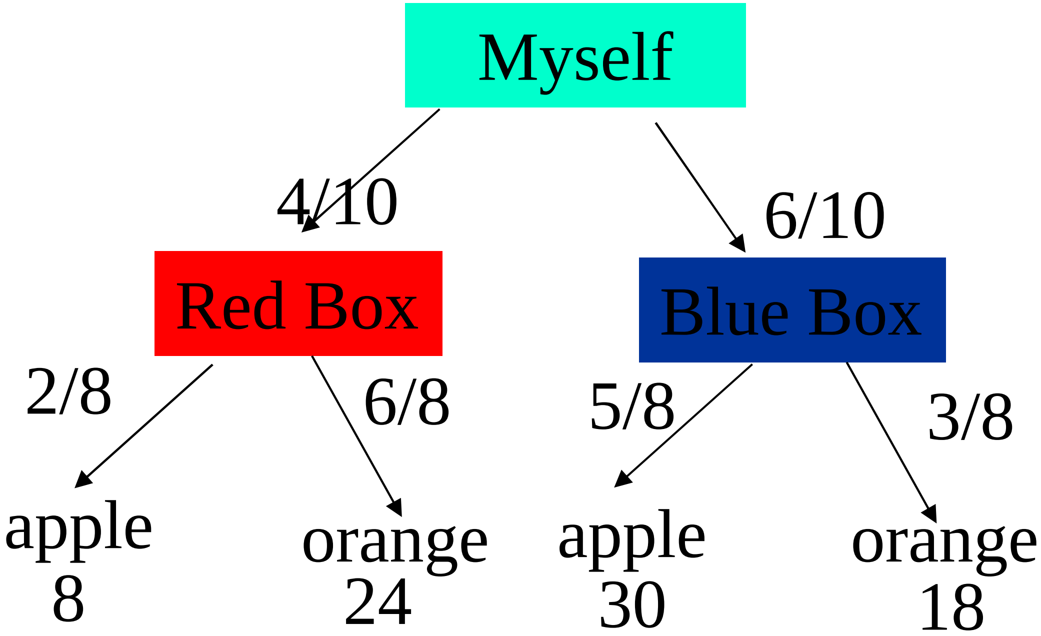


Supponiamo di conoscere P(X), probabilità di scelta del box, e la P(Y|X), probabilità di avere una mela (arancia) se scegliamo un certo box, possiamo determinare la probabilità assoluta (semplice) di scegliere un certo frutto, P(Y)?

Supponiamo di non conoscere P(X), probabilità di scelta del box, conosciamo la probabilità P(Y|X) e P(Y). Possiamo determinare P(X)?

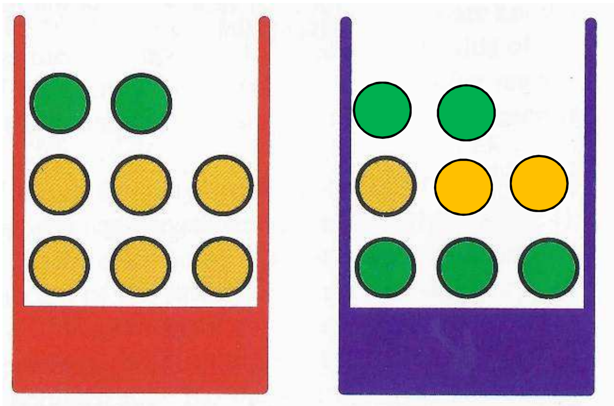
Determino P(Y)

$P(Y=apple) \neq (2+5) / 16 = 7/16$
$P(Y=orange) \neq (6+3)/16 = 9/16$



$P(Y=apple \mid X = blue) = 5/8$
$P(Y=orange \mid X = blue) = 3/8$

$P(Y=apple \mid X = red) = 2/8$
$P(Y=orange \mid X = red) = 6/8$



$P(Y=apple \mid X = blue) + P(Y=orange \mid X = blue) = 1$
$P(Y=apple \mid X = red) + P(Y=orange \mid X = red) = 1$

$P(Y=apple) = P(Y=apple \mid X = blue)\, P(X=blue) + P(Y=apple \mid X = red)\, P(X=red) = 5/8*6/10 + 2/8*4/10 = 38/80 \neq 7/16$

$P(Y=orange) = P(Y=orange \mid X = blue)\, P(X=blue) + P(Y=orange \mid X = red)\, P(X=red) = 3/8*6/10 + 6/8*4/10 = 42/80 \neq 9/16$

Determino P(Y) - tabella

	Red	Blue
Apple	4/10*2/8	6/10*5/8
Orange	4/10*6/8	6/10*3/8

y_j x_i

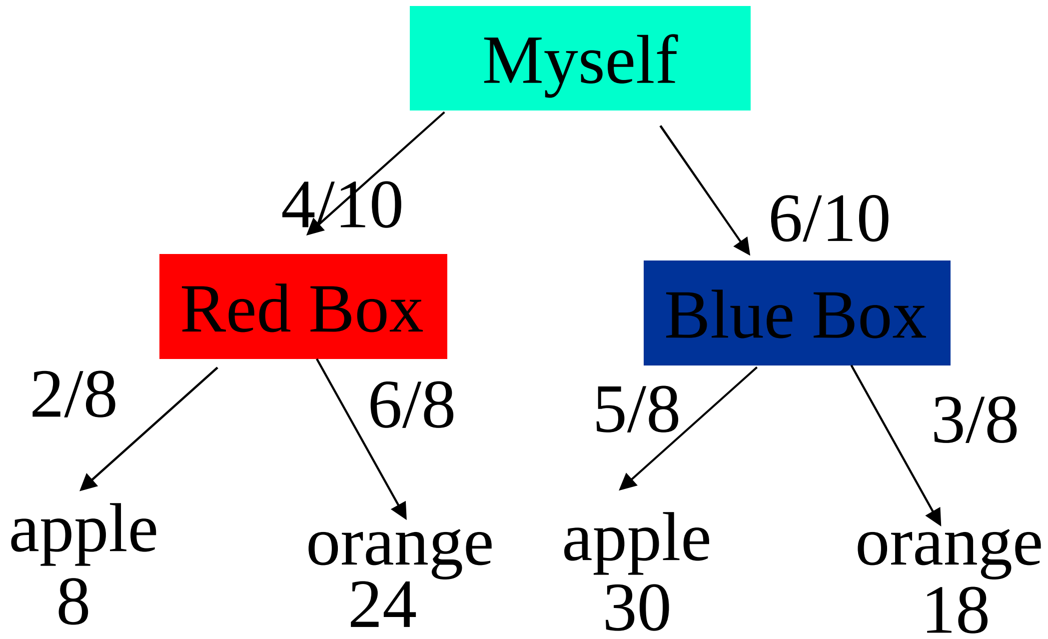
Sommo lungo x, marginalizzo x.

P(Y=apple) = 5/8*6/10 + 2/8*4/10 = 38/80
P(Y=orange) = 3/8*6/10 + 6/8*4/10 = 42/80

P(X=red) = 4/10*2/8 + 4/10*6/8 = 4/10
P(X=blue) = 6/10*5/8 + 6/10*3/8 = 6/10

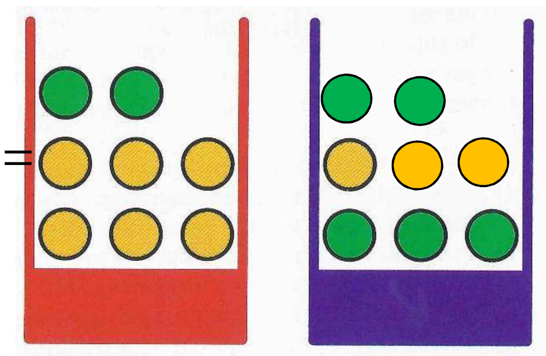
Determino P(X|Y)

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



P(X=red | Y=orange) = P(Y=orange|X=red) P(X=red)/P(Y=orange) =
(6/8*4/10) / (21/40) = 24/42 > 4/10

P(X=blue | Y=orange) = P(Y=orange|X=blue) P(X=blue)/P(Y=orange) =
(3/8*6/10) / (21/40) = 18/42 < 6/10

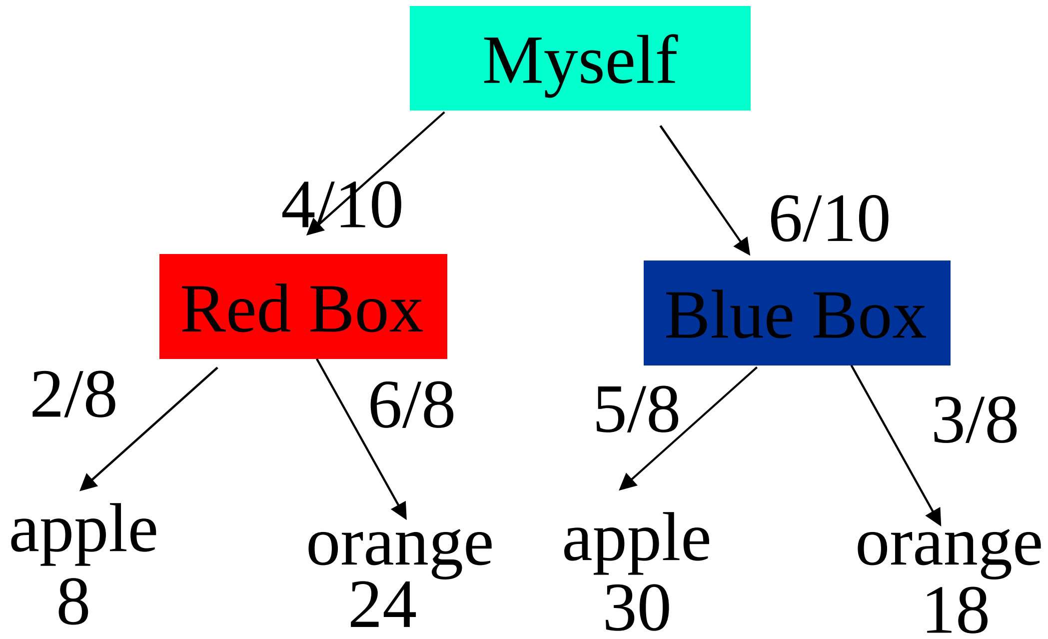


P(X=red | Y=apple) = P(Y=apple | X=red) P(X=red) / P(Y=apple) =
(2/8*4/10) / (19/40) = 8/38 << 4/10

P(X=blue | Y=apple) = P(Y=apple | X=blue) P(X=blue) / P(Y=apple) =
(5/8*6/10) / (19/40) = 30/38 >> 6/10

Interpretazione

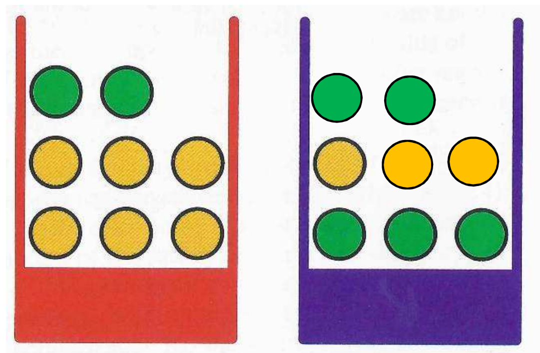
$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$



P(X=red | Y=apple) = 8/38 << 4/10

P(X=blue | Y=apple) = 30/38 >> 6/10

Correggo la probabilità a-priori, P(X) con le informazioni raccolte, P(Y), e ottengo una nuova valutazione della probabilità di X (che dipende da Y), P(X | Y) detta probabilità a-posteriori.

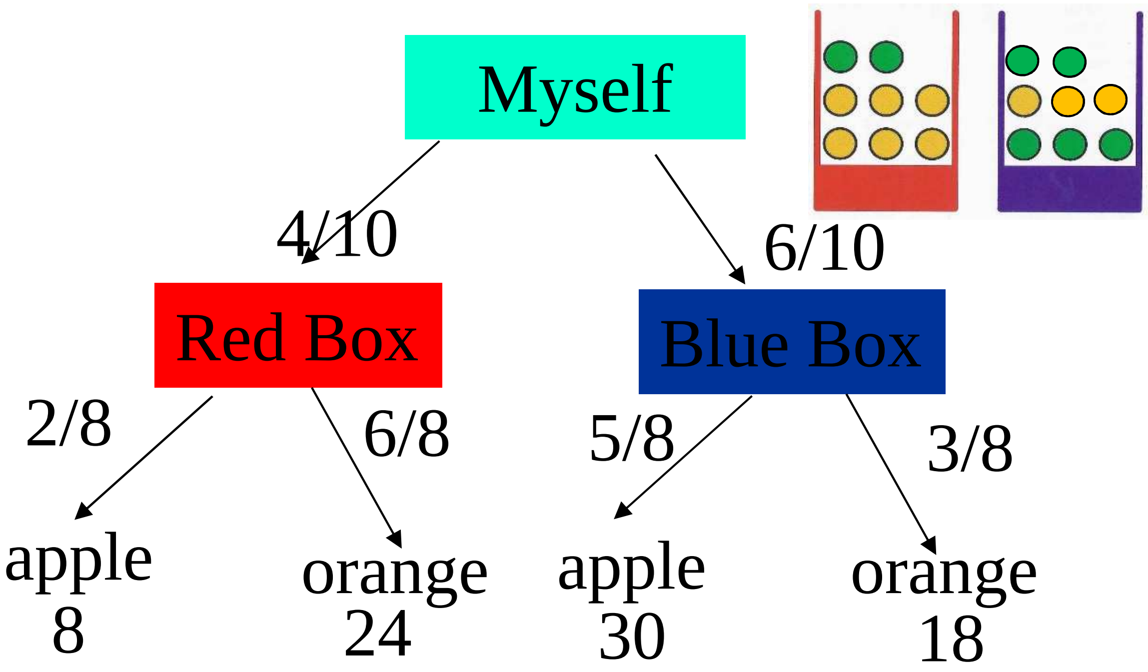


	Red	Blue	
Apple	4/10*2/8=8/80 – 8	6/10*5/8=30/80 – 30	y_j
Orange	4/10*6/8=24/80 - 24	6/10*3/8=18/80 - 18	

x_i

Importanza

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$



Questo è un tipico esempio di problema **inverso**.

Raccogliamo delle misure Y e vogliamo determinare da quale sistema (modello probabilistico) possono essere state generate.

Possiamo inserire delle informazioni statistiche (a-priori) su X, cioè sulla forma del modello (e.g. smoothness)

Overview

Probabilità semplice e condizionata

Esempio

Teorema di Bayes