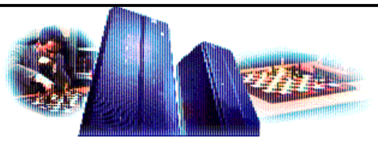
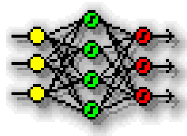




Sistemi Intelligenti
Introduzione al calcolo delle probabilità - I

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

Teorema di Bayes

Lo statistico

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 di incertezza legati al risultato delle azioni. 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.

La probabilità che si verifichi uno tra i vari casi possibili è sempre 1. Ovverosia la somma delle probabilità di tutti gli eventi (se mutuamente esclusivi) somma 1.

Supponiamo A = {a_1, a_2}

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

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

Probabilità totale

$$P(A) = \sum_{k=1}^{N} \left(\lim_{N\to\infty} \left(\frac{n_A = a_k}{N} \right) \right)$$

Altri modi di definire la probabilità

- **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.

Probabilità

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!! La conoscenza a-priori e' ricavata dall'analisi di tutti gli altri pazienti già osservati o su una conoscenza oggettiva o soggettiva dell'evento.

La probabilità ci consente di trattare le diverse possibilità di un evento.

E ci consente di associare agli eventi un grado di credenza, prima che si verifichino.

Problematiche associate alla probabilità

Problema della visione frequentista:

- **Omogeneità del campione** (classe di riferimento). Come posso effettuare la media di eventi in modo "sicuro"?
- **Limitatezza del campione**

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 (mal di denti del paziente).

Combinazione di probabilità con i connettivi logici (probabilità congiunta)

Qual'è la probabilità che la proposizione: "E' uscito 12" tirando due dadi si avveri?

Probabilità congiunta di due eventi indipendenti

$P(N=12) = P(dado_1 = 6 \text{ AND } dado_2 = 6) = P(dado_1 = 6, dado_2 = 6) = P(dado_1 = 6)\, P(dado_2 = 6) = 1/36.$

Nel caso di **indipendenza**, la probabilità di A e B è data dal **prodotto** delle probabilità:
$P(X = A \text{ AND } Y = B) = P(X=A)\, P(Y=B) = P(X=6)\, P(Y=6)$

Combinazione di probabilità con i connettivi logici (probabilità congiunta)

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

$P(N=11) = P(dado_1 = 5 \text{ AND } dado_2 = 6) \text{ OR } P(dado_1 = 6 \text{ AND } dado_2 = 5) = 1/6*1/6 + 1/6*1/6 = 2/36$

Probabilità congiunta di due eventi indipendenti

Probabilità Totale

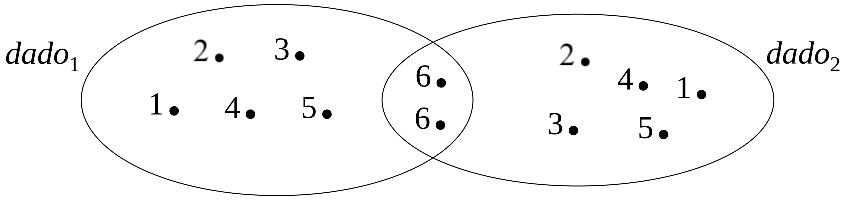
$P(N=11) = P(X=5 \text{ AND } Y = 6) \text{ OR } P(X=6 \text{ AND } Y = 5)$

Combinazione di probabilità con i connettivi logici (probabilità congiunta)

Qual'è la probabilità che la proposizione: "E' uscito almeno un 6 si avveri?

$P(N=6) = P(dado_1 = 6 \text{ OR } dado_2 = 6) - P(dado_1 = 6 \text{ AND } dado_2 = 6) = 1/6+1/6 - 1/6*1/6 = 11/36!$

Non voglio contare due volte la probabilità di ottenere 6.



In generale: P(X = A OR Y = B) = P(X=A) + P(Y=B) – P(X=A AND Y=B).

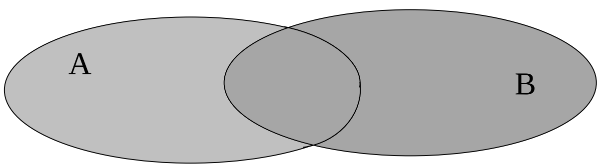
Se gli eventi sono disgiunti, appartengono a due insiemi diversi:
P(X = A OR Y = B) = P(X=A) + P(Y=B) – 0. Probabilità congiunta.

Probabilità incondizionate o a-priori. Non richiedono o dipendono da altre informazioni.

Probabilità di 2 variabili

P(X = A **OR** Y = B)=(X=A)+P(Y=B) **– P(X=A AND Y=B)** =>
se sono indipendenti = P(X=A)+P(X=B)

P(X = A **AND** Y = B)= P(X=A) + P(Y=B) **– P(X=A OR Y=B)** =>
se sono indipendenti = P(X=A) P(Y=B)



Dipendenza

- Cosa succede se X e Y non sono indipendenti, ma Y dipende da X e X è indipendente? P(X) => P(Y)?

- Cosa succede se esiste una relazione funzionale tra X e Y, dove X è indipendente, mentre Y dipende da quello che fa la variabile X?
  - "Y = f(X)" nel dominio analitico.
  - "P(Y|X)" nel dominio statistico.

Dipendenza tra probabilità

Supponiamo ora che il primo dado abbia mostrato 5. Abbiamo un'informazione prima di lanciare il secondo dado. Affinchè N = 11, occorre che il secondo dato mostri 6.
Questo è diverso dal chiedere quale sia la probabilità che lanciati due dadi si ottenga 11. C'e' una condizioni prima di lanciare il secondo dato (a-priori).

$P(N=11 \mid Dado_1 = 5) = 1/6$ > P(N=11) lanciando 2 dadi. Abbiamo un'incertezza minore.

Probabilità condizionata.

P(Y=A | X=B)

Un agente cerca di raccogliere più informazioni possibili per diradare l'incertezza e formulare quindi una soluzione più certa. I problemi sono descrivibili con probabilità condizionate.

La probabilità condizionata stabilisce una precedenza, una corrispondenza, una dipendenza funzionale tra X e Y. Dato X, determino Y.

Relazione tra probabilità condizionata e congiunta

Nel caso dei dadi, quando c'e' dipendenza: $P(N = 11 \mid Dado_1 = 5) = 1/6$

$P(a \text{ AND } b) = P(a \mid b) * P(b)$ — $P(a \text{ AND } b)$ è probabilità congiunta; $P(a \mid b)$ è probabilità condizionata

Se a e b sono indipendenti: P(a AND b) = P(a)P(b), probabilità congiunta di eventi indipendenti:

$P(N=11 \text{ AND } Dado_1 = 5) = P(N = 11 \mid Dado_1 = 5) * P(Dado_1 = 5) = (1/6) * (1/6) = 1/36)$

Consideriamo prima b e poi a.

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

Possiamo anche scrivere:

$P(N = 11 \mid Dado_1 = 5) = P(N=11 \text{ AND } Dado_1 = 5) / P(Dado_1 = 5) = (1/36) / (1/6) = 1/6$

Probabilità condizionata e semplice

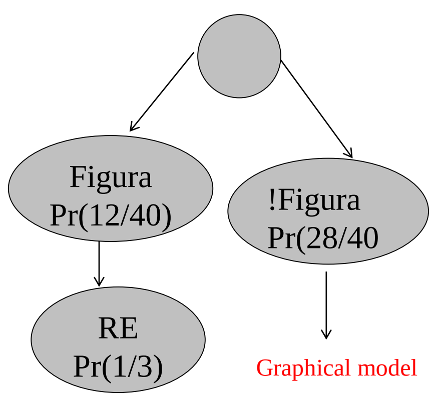
Consideriamo un mazzo di 40 carte:

- vogliamo valutare quale sia la probabilità che una carta estratta a caso sia un re (probabilità semplice)
- vogliamo valutare quale sia la probabilità che una carta estratta a caso sia un re, sapendo di avere estratto una figura (probabilità condizionata)

P(Y) = probabilità che sia un re
P(X) = probabilità che sia una figura

Le figure sono 12 in un mazzo di 40 carte
P(Y | X) = P(re | figura) =
1/12+1/12+1/12+1/12 = 1/3

P(Y) = P(Y/X) P(X) = 1/3 12/40 = 4/40



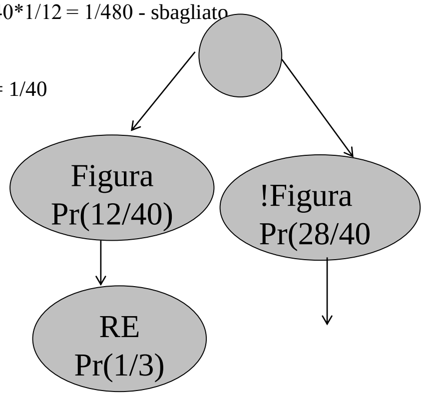
Graphical model

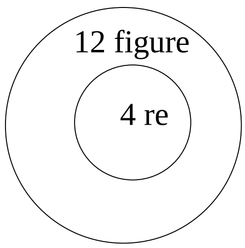
Probabilità congiunta - I

Probabilità di avere un re di cuori o un re di quadri
P ((Y= re_cuori) OR (Y = re_quadri)) = P(Y = re_quadri) + P(Y = re_cuori) = 2/40

Probabilità di avere un re di cuori **e** una figura (non sono indipendenti!)
P((Y = re_cuori) AND (Y = figura)) ≠ P(Y=re_cuori)P(figura) =1/40*1/12 = 1/480 - sbagliato

P((Y = re_cuori) AND (Y = figura)) = P(Y=re_cuori) + P(figura)
– **P(Y=re_cuori) OR (Y=figura))** = 1/40 + 1/12 – 1/12 = 1/40



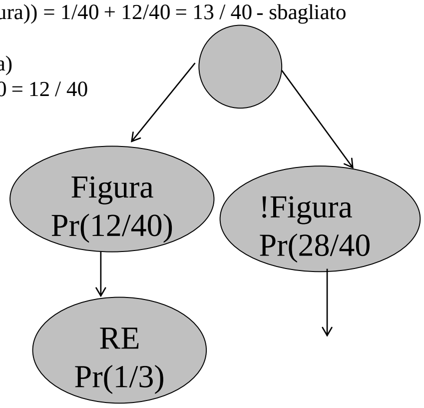


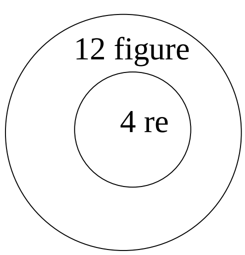
Probabilità congiunta - II

Probabilità di avere un re di cuori o un re di quadri
P ((Y= re_cuori) OR (Y = re_quadri)) = P(Y = re_quadri) + P(Y = re_cuori) = 2/40

Probabilità di avere un re di cuori **o** una figura (non sono indipendenti!)
P((Y = re_cuori) OR (Y = figura)) ≠ P((Y = re_cuori) + P(Y = figura)) = 1/40 + 12/40 = 13 / 40 - sbagliato

P((Y = re_cuori) OR (Y = figura)) = P(Y = re_cuori) + P(Y = figura)
– P(Y=re_Cuori AND Y = figura) = 1/40 + 12/40 – 1/40 = 12 / 40





Inferenza statistica

- Calcolo della probabilità di un evento, a partire dall'informazione collezionata sperimentalmente (visione frequentista).
- Consideriamo tre variabili binarie: Mal di denti, Carie, Cavità in dente, e le probabilità congiunte (stimate dal dentista in base alla sua esperienza passata):

	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) = P(carie) + P(mal di denti) – P(carie AND mal di denti) = 0,108 + 0,012 + 0,072 + 0,008 + 0,108 + 0,012 + 0,016 + 0,064 - (0,108 + 0,012) = 0,2 + 0,2 – 0,12 = 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 AND mal di denti) = P(carie) + P(mal di denti) – P(carie OR mal di denti) = 0,108 + 0,012 + 0,072 + 0,008 + 0,108 + 0,012 + 0,016 + 0,064 – (0.28) = 0.2 + 0.2 – 0.28 = 0.12

Marginalizzazione

	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.

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

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 (assolute):

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
P(carie | mal di denti)	carie	0,54	0,06	0,09	0,01
P(carie | !mal di denti)	!carie	0,08	0,32	0,18	0,72

Probabilità condizionate al mal di denti:
P(carie| mal di denti) = P(carie AND mal di denti) / P(mal di denti) - Divido per 0.2

P(carie | !mal di denti)
Divido per 0.8

Probabilità condizionata

	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 condizionata
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

Probabilità congiunta
P(carie AND 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

Overview

Probabilità semplice e condizionata

Teorema di Bayes

Teorema di Bayes (1701-1761)

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

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

X = causa Y = effetto

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

Thomas Bayes

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)

Esempio A - I

Software in Scilab available!

In una città lavorano due compagnie di taxi:
blu e verde: $X = \{T_{blu}, T_{verde}\}$

con una Distribuzione di 85% di taxi verdi e 15% di taxi blu.
P(Taxi = verde) = 0.85
P(Taxi = blu) = 0.15

Succede un incidente in cui è coinvolto un taxi.

Un testimone dichiara che il taxi era blu P(Y=blu). Era sera e l'affidabilità del testimone è stata valutata dell'80% (è una probabilità condizionata):
P(Y=blu | X =blu) = P(Y=verde | X = verde) = 0.8

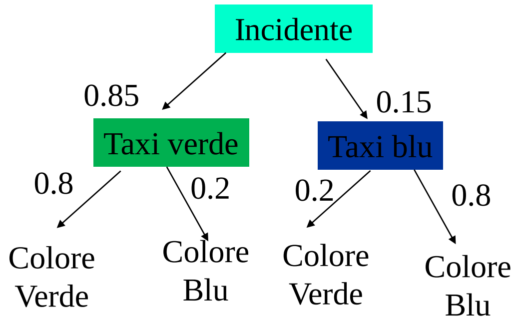
Qual è la probabilità che il taxi fosse effettivamente blu?

Non è l'80%!

Esempio A - II

X = {Taxi = blu, Taxi = verde }
"Causa"

Y = {Colore = blu, Colore = verde}
"Effetto"



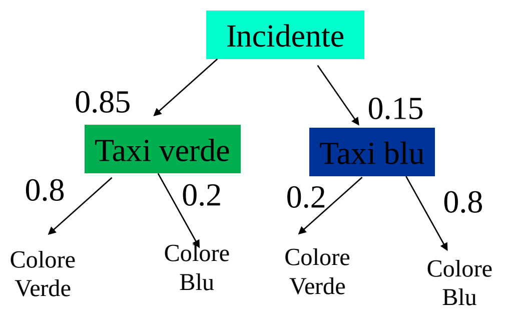
P(Taxi = blu) = Probabilità a-priori = 0.15
P(Colore = blu | Taxi = blu) = P(Colore = verde | Taxi = verde) – Affidabilità del testimone - Probabilità condizionata = 0.8

Come combino queste informazioni per ottenere una stima sulla probabilità che il taxi dell'incidente sia effettivamente blu?

Esempio A - III

X = {Taxi = blu, Taxi = verde }
"Causa"

Y = {Colore = blu, Colore = verde}
"Effetto"



Inverto la relazione tra causa ed effetto applicando Bayes:

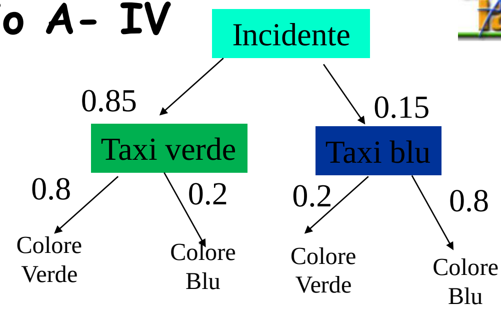
$$\boxed{P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}}$$

P(Taxi = blu | Colore = blu) =
P(Colore = blu | Taxi = blu)P(Taxi = blu) / P(Colore = blu)

Esempio A- IV

X = {Taxi = blu, Taxi = verde }
"Causa"

Y = {Colore = blu, Colore = verde}
"Effetto"



P(Colore = blu) = {Probabilità marginale di Y (probabilità semplice)} =
P(Colore = blu | Taxi = blu)P(Taxi = blu) +
P(Colore = blu | Taxi = verde)P(Taxi = verde) = 0.8*0.15 + 0.2*0.85 = 0.29

P(Colore = blu | Taxi = blu) = 0.8 {Affidabilità del testimone}
P(Colore = verde | Taxi = verde) = 0.8 {Affidabilità del testimone}

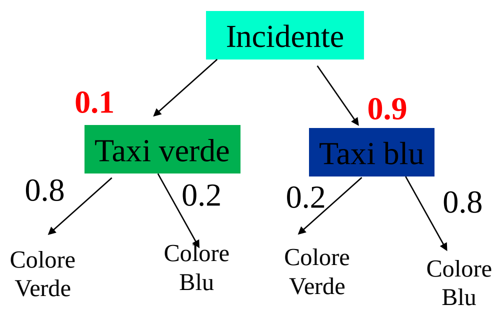
P(Taxi = blu) = 0.15 {Probabilità a-priori che si incontri un taxi blu}
P(Taxi = verde) = 0.85 {Probabilità a-priori che si incontri un taxi verde}

P(Taxi = blu | Colore = blu) =
P(Colore=blu | Taxi=blu)P(Taxi=blu) / P(Colore=blu) = 0.8*0.15 / 0.29 = 0.41 **0.15 < 0.41 << 0.8!!**
Pesano anche gli "errori" commessi quando il testimone vede un blu ma era verde!

Esempio A - V

X = {Taxi = blu, Taxi = verde }
"Causa" – Molti più taxi blu.
P(Taxi = verde) = 0.1
P(Taxi = blu) = 0.9

Y = {Colore = blu, Colore = verde}
"Effetto"



P(Colore = blu | Taxi = blu) = P(Colore = blu | Taxi = blu)P(Taxi = blu) / P(Colore = blu)

P(Colore = blu) = {Probabilità marginale di Y (probabilità semplice)} =
P(Colore = blu | Taxi = blu)P(Taxi = blu) + P(Colore = blu | Taxi = verde)P(Taxi = verde) =
0.8*0.9 + 0.2*0.1 = 0.74

P(Taxi = blu | Colore = blu) = P(Colore = blu | Taxi = blu)P(Taxi = blu) / P(Colore = blu) =
0.8*0.9 / 0.74 = 0.97
Testimonianza molto affidabile in questo caso!

Esempio B - I

Lo strumento principe per lo screaning per il tumore al seno è la radiografia (mammografia).

Definiamo X la situazione della donna: X={sana, malata}
Definiamo Y l'esito della mammografia: Y={positiva, negativa}

La percentuale di donne malate sulla popolazione è dell'1%.
P(X) = 0.01 – probabilità a-priori.

La sensitività della mammografia è intorno al 90%:

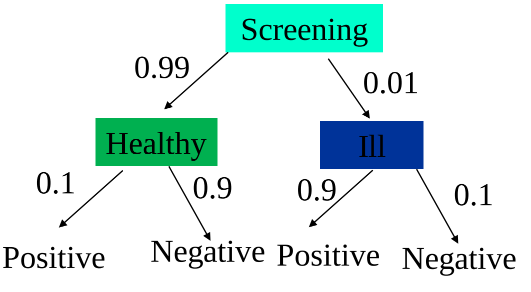
$$\text{sensitività} = \frac{n_{positive}}{N_{ill}}$$ => **P(Y=positive | X=ill (positive))**

La specificità della mammografia è anch'essa intorno al 90%:

$$\text{specificità} = \frac{n_{negative}}{N_{healthy}}$$ => **P(Y=negative | X=healthy (negative))**

Esempio B - II

X = {Healthy, Ill}
Y = {Positive, Negative}



P(Y=Positive | X=Ill)*P(X=Ill) = 0.9 * 0.01 = 0.009
P(Y=Positive | X=Healthy)*P(X=Healthy) = 0.1*0.99 = 0.099

P(Y=Positive) = P(Y=Positive | X=Ill)*P(X=Ill) + P(Y=Positive | X=Healthy)*P(X=Healthy) =
0.009 + 0.099 = 0.108 (probabilità marginale di Y)

10.8% di probabilità di avere un esame positivo a fronte di uno 0.01% di donne malate!
Solo lo 0,9% proviene da donne effettivamente malate, le altre sono false positive!

Esempio B - III

10.8% di probabilità di avere un esame positivo a fronte di uno 0.01% di donne malate!
Solo lo 0,9% circa proviene da donne effettivamente malate, le altre sono false positive!

Qual'è la probabilità che una donna sia veramente malata se il test risulta positivo?

Applichiamo Bayes P(X=Ill | Y=Positive) - PPV (Positive Predictive Value)

P(X=Ill | Y=Positive) = P(Y=Positive | X=Ill)P(X=Ill) / P(Y=Positive) =
0.09 / 0.108 = 0.083 (8.3%)

Solo 8.3% delle donne con mammografia positiva sono effettivamente ammalate.

Analizzando la formula del teorema di Bayes, dove ha senso investire per ottenere un rendimento delle screening maggiore?

Valutazione delle prestazioni dei test binari

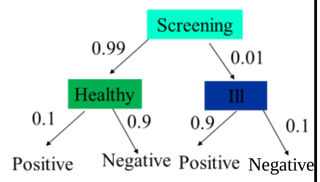
	Malati (positivi)	Sani (negativi)
Classifico Positivi	Veri +	Falsi +
Classifico Negativi	Falsi -	Veri -

Sensitività: $\frac{Veri+}{N_{pos}} = \frac{Veri+}{(Veri+) + (Falsi-)} = \frac{Veri+}{Num\ malati}$

Specificità: $\frac{Veri-}{N_{neg}} = \frac{Veri-}{(Veri-) + (Falsi+)} = \frac{Veri-}{Num\ sani}$

Positive predictive value: $\frac{Veri+}{N_{class\_pos}} = \frac{Veri+}{(Veri+) + (Falsi+)}$

Negative predictive value: $\frac{Veri-}{N_{class\ neg}} = \frac{Veri-}{(Veri-) + (Falsi-)}$

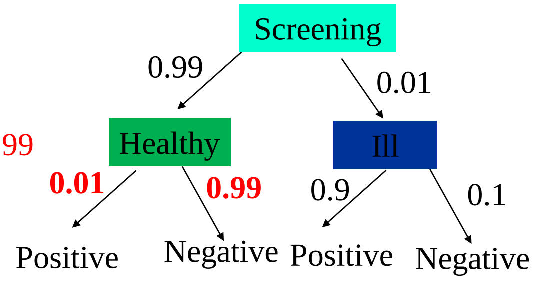


Dove conviene investire?

Investiamo sulla specificità del test

X = {Healthy, Ill}
Y = {Positive, Negative}

P(Y=Negative | X=Healthy) = 0.99



P(Y=Positive | X=Ill) = 0.9 * 0.01 = 0.009

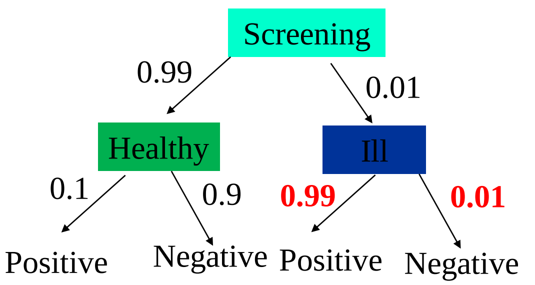
P(Y=Positive) = P(Y=Positive | X=Ill) + P(Y=Positive | X=Healthy) =
0.009 + 0.99*0.01 = 0.0189

P(X=Ill | Y=Positive) = P(Y=Positive | X=Ill)P(X=Ill) / P(Y=Positive) =
= 0.009 / 0.0189 = 0.476 = 47,6% >> 8.3%

Investiamo sulla Sensitività del test

X = {Healthy, Ill}
Y = {Positive, Negative}

P(Y=Positive | X= Ill) = 0.99



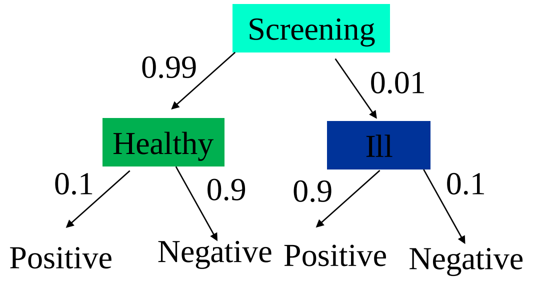
P(Y=Positive | X=Ill) = 0.99 * 0.01 = 0.0099

P(Y=Positive) = P(Y=Positive | X=Ill) + P(Y=Positive | X=Healthy) =
0.0099 + 0.99*0.1 = 0. 1098

P(X=Ill | Y=Positive) = P(Y=Positive | X=Ill)P(X=Ill) / P(Y=Positive) =
= 0.0099 / 0.1098 = 0.09 = 9% > 8.3%.

Esempio B - IV

X = {Healthy, Ill}
Y = {Positive, Negative}



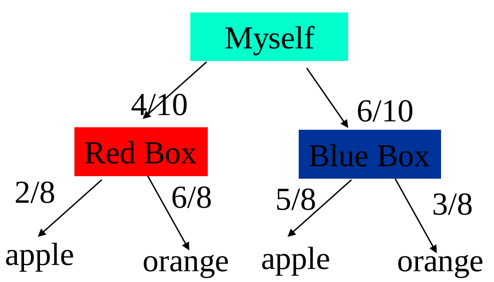
Falsi negativi?
P(X = Ill | Y = Negative)?

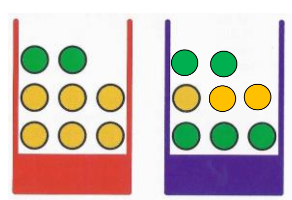
P(Y=Negative | X=Ill) = 0.1 * 0.01 = 0.001

P(Y=Negative) = P(Y=Negative | X=Ill) + P(Y=Positive | X=Healthy) =
0.001 + 0.99*0.9 = 0.891

P(X=Ill | Y=Negative) = P(Y=Negative | X=Ill)P(X=Ill) / P(Y=Negative) = 0.001 / 0.891 = 0.11%
Una donna ogni mille non viene diagnosticata!

Esempio C - I





1) Supponiamo di conoscere P(X), probabilità di scelta del box,
P(blu) = 0.6
P(rosso) = 0.4
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)?

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

1) Determino P(Y = Apple)

P(Y=apple | X = blue) = 5/8 - nota
P(Y=apple | X = red) = 2/8 - nota

P(Y=apple) ≠ (2+5) / 8 = 7/8

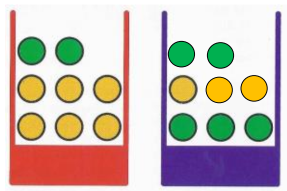
Myself

4/10 — Red Box; 6/10 — Blue Box

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

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

P(Y=apple | X = blue) + P(Y=orange | X = blue) = 1
P(Y=apple | X = red) + P(Y=orange | X = red) = 1



P(Y=apple) = P(Y=apple | X = blue) P(X=blue) +
P(Y=apple | X = red) P(X=red) = 5/8*6/10 + 2/8*4/10 = 38/80 ≠ 7/8

1) Determino P(Y = orange)

P(Y=apple | X = blue) = 5/8 - nota
P(Y=orange | X = blue) = 3/8 - nota

P(Y=apple | X = red) = 2/8 - nota
P(Y=orange | X = red) = 6/8 – nota

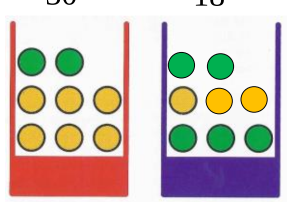
P(Y=apple) ≠ (2+5) / 8 = 7/8
P(Y=orange) ≠ (6+3)/8 = 9/8

Myself

4/10 — Red Box; 6/10 — Blue Box

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

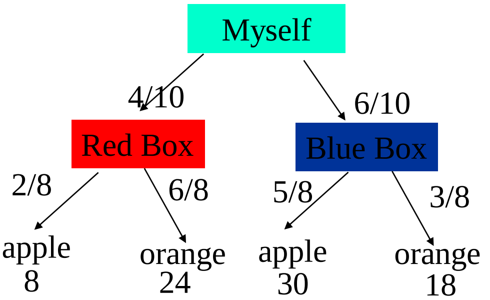
Blue Box: 5/8 apple 30; 3/8 orange 18



P(Y=orange) = P(Y=orange | X = blue) P(X=blue) +
P(Y=orange | X = red) P(X=red) = 3/8*6/10 + 6/8*4/10 = 42/80 ≠ 9/8
P(Y=apple) = 38/80
P(Y=apple) + P(Y=orange) = 1

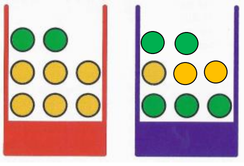
Determino P(X=red | Y)

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



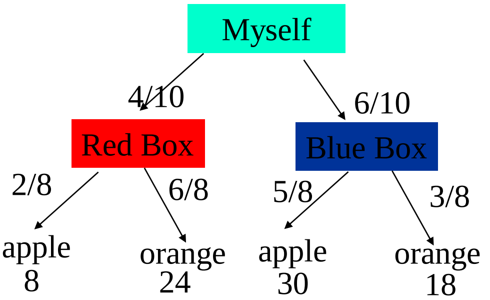
P(X=red | Y=orange) = P(Y=orange|X=red) P(X=red)/P(Y=orange) =
(6/8*4/10) / (21/40) = 24/42 > 4/10

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



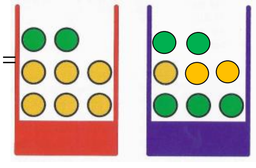
Determino P(X | Y)

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



P(X=red | Y=orange) = 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

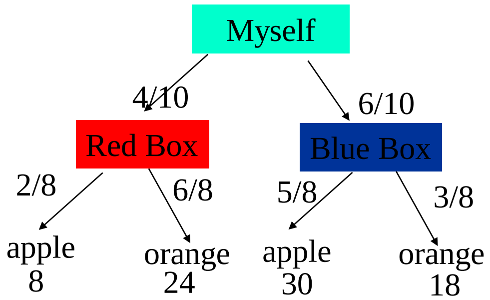
P(X=red | Y=apple) = 8/38 << 4/10
P(X=blue | Y=apple) = P(Y=apple | X=blue) P(X=blue) / P(Y=apple) =
(5/8*6/10) / (19/40) = 30/38 >> 6/10

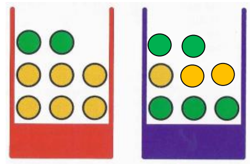


Interpretazione

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

P(X=red | Y=apple) = 8/38 << 4/10
P(X=blue | Y=apple) = 30/38 >> 6/10





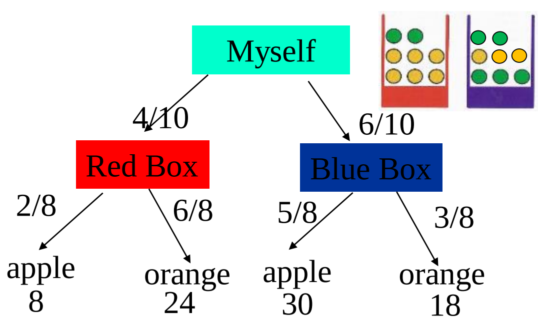
Correggo la probabilità a-priori, P(X) con le informazioni raccolte, P(Y), e ottengo una nuova valutazione della probabilità di X (che dipende da Y), P(X | Y) detta probabilità a-posteriori.

	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

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



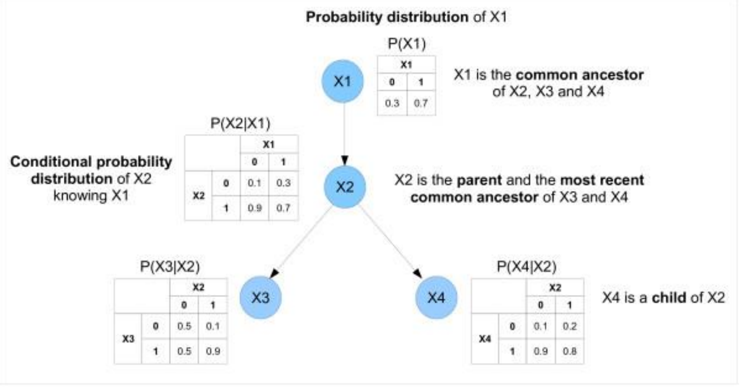
Questo è un tipico esempio di problema **inverso**.

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)

Graphical models

A **graphical model** o **modello probabilistico su grafo (PGM)** è un modello probabilistico che evidenzia le dipendenze tra le variabili randomiche (può evolvere eventualmente in un albero). Viene utilizzato nell'inferenza statistica.



Estensione a più variabili

Sostituisco un'espressione logica a una variabile:

$P(X|\ Y_1;Y_2)$ if ($P(Y_1) = y_1$ and $P(Y_2) = y_2$) then $P(X) = x$

$Z = Y_1$ *and* Y

	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

Dalla tabella delle probabilità congiunte ricaviamo (Z = mal di denti AND cavità):

P(carie; Z) = P(carie AND (mal di denti AND cavità)) = 0,108

P(carie | Z) = P(carie | (mal di denti AND cavità)) =
P(carie AND (mal di denti AND cavità)) / P(mal di denti AND cavità) =
0,108 / (0,016 + 0,108) = 0,108 / 0,124 = 0,871

Estensione a più variabili

$P(X|Y_1;Y_2)$ if ($P(Y_1) = y_1$ and $P(Y_2) = y_2$) then $P(X) = x$

$Z = Y_1$ *and* Y_2

	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 | Z) = P(carie | (mal di denti AND cavità)) =
P(carie AND (mal di denti AND cavità)) / P(mal di denti AND cavità) =
0,108 / (0,016 + 0,108) = 0,108 / 0,124 = 0,871

P(mal di denti AND cavità | carie) = P((mal di denti AND cavità AND carie) / P(carie) =
0,108 / 0,2 = 0,54

Carie, mal di denti e cavità non sono indipendenti, come trattiamo l'AND?

Estensione a più variabili

$P(X|Y_1;Y_2)$ if ($P(Y_1) = y_1$ and $P(Y_2) = y_2$) then $P(X) = x$

$Z = Y_1$ *and* Y_2

	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

Carie, mal di denti e cavità non sono indipendenti, come trattiamo l'AND?

P(carie AND mal di denti) = P(carie) + P(mal di denti) – P(carie OR mal di denti) = 0,108 + 0,012 + 0,072 + 0,008 + 0,108 + 0,012 + 0,016 + 0,064 – (0.28) = 0.2 + 0.2 – 0.28 = 0.12

Come lo possiamo trattare invece se sono indipendenti?

P(X AND Y) = P(X)P(Y)

Naïve Bayes

Sostituiamo un'espressione a una variabile logica:
$P(X \mid (Y_1 \text{ and } Y_2))*P(Y_1 \text{ and } Y_2) = P(Y_1 \text{ and } Y_2 \mid X) * P(X)$ Bayes

Introduciamo un'altra ipotesi. Cosa succede se Y_1 e Y_2 sono indipendenti? Dipendono entrambe da X ma non dipendono tra di loro.

Sono cioè **condizionatamente indipendenti**, cioè vale che:
$\mathbf{P((Y_1 \text{ and } Y_2) \mid X) * P(X) = (P(Y_1 \mid X) * P(Y_2 \mid X)) * P(X)}$

In questo caso viene semplificato il calcolo dell'AND, che viene calcolato come prodotto delle probabilità.

Modello Naive Bayes Gli effetti sono supposti indipendenti tra loro e dipendono da una stessa causa

In generale: P(Causa | Effetto_1 and Effetto_2 and ... Effetto_N) = $\boxed{\prod_{i=1}^{N} P(Effetto_i \mid Causa)}$

Riepilogo

$$\boxed{P(X|Y) = \frac{P(Y \mid X)P(X)}{P(Y)} = \frac{P(X,Y)}{P(Y)}}$$ Teorema di Bayes

Lega probabilità condizionate, congiunte, semplici (marginali)

Consente di inferire la probabilità di un evento causa, X, a partire dalla probabilità associata alla frequenza di una certa misura, effetto, P(Y), dalla frequenza relativa dell'evento associato alla misura, P(Y), e dalla probabilità nota a-priori, P(X), della causa.

La probabilità P(X|Y) viene per questo detta probabilità a-posteriori ed è una probabilità condizionata.

Viene utilizzata nei problemi inversi.

Overview

Probabilità semplice e condizionata

Teorema di Bayes