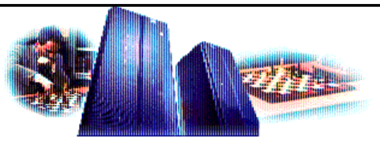
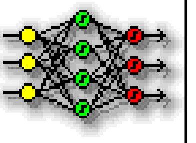




Introduzione al calcolo delle probabilità - II

Alberto Borghese

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

Overview

Esempi sul teorema di Bayes

Densità di probabilità

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!!

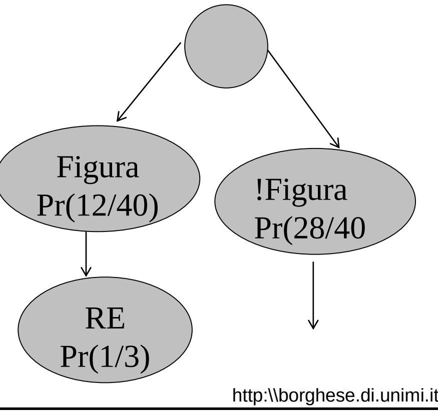
Probabilità condizionata

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

P(Y | X) = 1/3

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



Esempio probabilità condizionata

P(X, Y)

	mal di denti	!mal di denti
carie	0,12	0,08
!carie	0,08	0,72

P(Y | X)

	mal di denti	!mal di denti
carie	0,6	0,4
!carie	0,1	0,9

P(X | Y)

	mal di denti	!mal di denti
carie	0,6	0,1
!carie	0,4	0,9

Teorema di Bayes

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

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

X = causa Y = effetto

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

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)

Esempio teorema di Bayes

P(X, Y)

	mal di denti	!mal di denti
carie	0,12	0,08
!carie	0,08	0,72

P(Y | X)

	mal di denti	!mal di denti
carie	0,6	0,4
!carie	0,1	0,9

P(X | Y)

	mal di denti	!mal di denti
carie	0,6	0,1
!carie	0,4	0,9

Calcoliamo P(Carie | ! Mal di denti)

$$0.1 = P(X \mid Y) = \frac{P(Y \mid X) * P(X)}{P(Y)} = 0.4 * 0.2 / 0.8$$

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

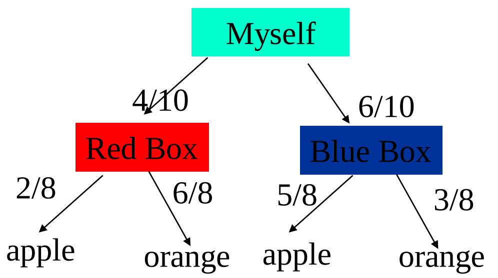
This formulation is interesting as far as we have sufficient knowledge and it is based on empirical evidence. Any doctor knows that 0,14% is the incidence of menigitis with stiff neck. However, what happens if there is an epidemy of meningitis?

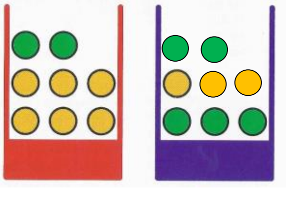
A doctor that uses knowledge not from experience would not change the odds.

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)?

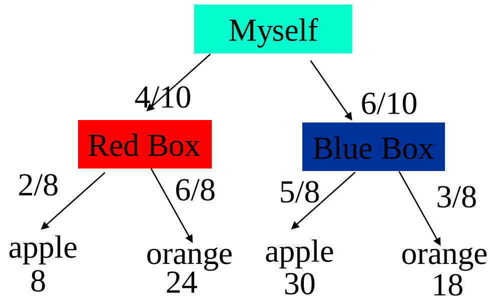
Supponiamo di non conoscere P(X), probabilità di scelta del box, conosciamo la probabilità P(Y|X) e P(Y). Possiamo determinare P(X)?

Determino P(Y)

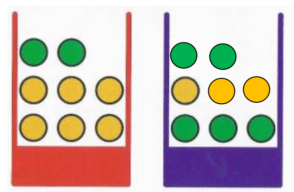
P(Y=apple) ≠ (2+5) / 16 = 7/16
P(Y=orange) ≠ (6+3)/16 = 9/16

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

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



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 ≠ 7/16
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 ≠ 9/16

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

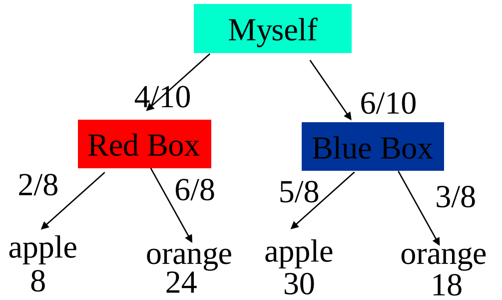
Sommo lungo x, marginalizzo x.

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

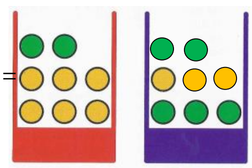
Determino P(X|Y)

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



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



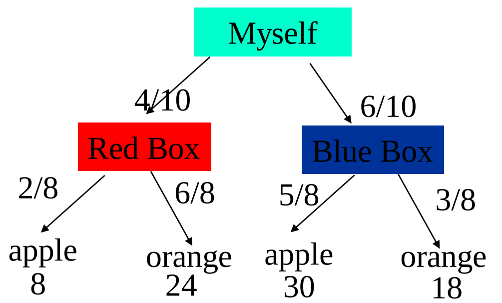
P(X=red | Y=apple) = P(Y=apple | X=red) P(X=red) / P(Y=apple) =
(2/8*4/10) / (19/40) = 8/38 << 4/10

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

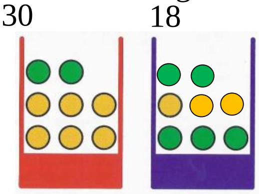
Interpretazione

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

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



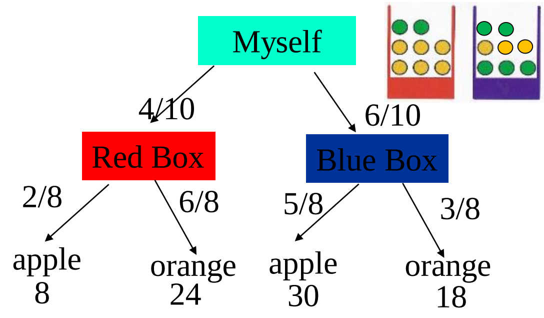
Correggo la probabilità a-priori, P(X) con le informazioni raccolte, P(Y), e ottengo una nuova valutazione della probabilità di X (che dipende da Y), P(X | Y) detta probabilità a-posteriori.



	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)

Esempio I

In una città lavorano due compagnie di taxi:
blue e verde: X = {T_{blu}, T_{verde}} con una
Distribuzione di 85% di taxi verdi e 15% di taxi blu.

Succede un incidente in cui è coinvolto un taxi.

Un testimone dichiara che il taxi era blu. Era sera e l'affidabilità del testimone è stata valutata dell'80%.

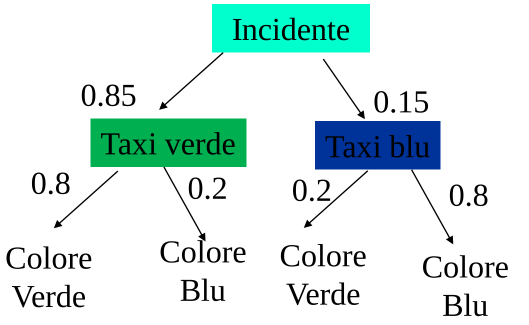
Qual è la probabilità che il taxi fosse effettivamente blu?

Non è l'80%!

Esempio - I

$X = \{T_{blu}, T_{verde}\}$
"Causa"

$Y = \{C_{blu}, C_{verde}\}$
"Effetto"

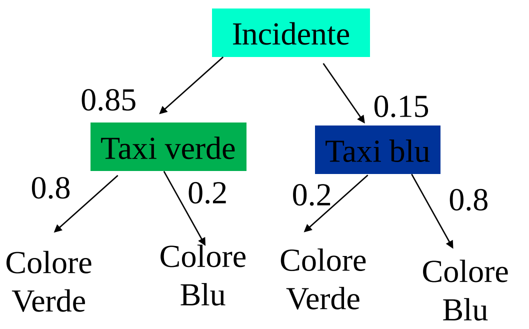


$$P(X=T_{blue}|Y_{blue}) = P(Y_{blue}|X_{blue})P(X_{blue}) / P(Y_{blue})$$

$P(X_{blue})$ = Probabilità a-priori = 0.15
$P(Y_{blue}|X_{blue})$ = Probabilità condizionata = 0.8

Esempio - I

$X = \{T_{blu}, T_{verde}\}$
$Y = \{C_{blu}, C_{verde}\}$



Inverto la relazione tra causa ed effetto applicando Bayes:
$P(X=T_{blu} \mid Y_{blu}) = P(Y_{blu}|X_{blu})P(X_{blu}) / P(Y_{blu})$

$P(Y_{blu})$ = Probabilità marginale di Y (probabilità semplice) =
$P(Y_{blu}|X_{blu})P(X_{blu}) + P(Y_{blu}|X_{verde})P(X_{verde}) = 0.8*0.15 + 0.2*0.85 = 0.29 > 0.15$!!

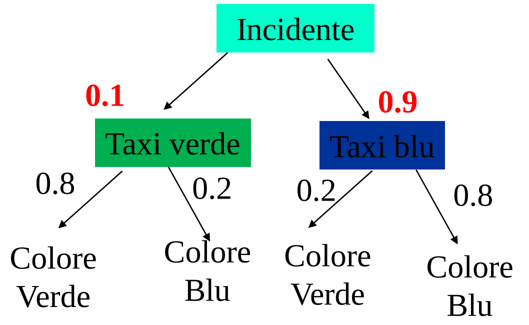
$P(X=T_{blu}|Y_{blu}) = P(Y_{blu}|X_{blu})P(X_{blu}) / P(Y_{blu}) = 0.8*0.15 / 0.29 = 0.41$
Pesano anche gli "errori" commessi quando il testimone vede un taxi verde!

Esempio - I

Cambio la distribuzione dei taxi

$X = \{T_{blu}, T_{verde}\}$
$Y = \{C_{blu}, C_{verde}\}$

$P(X{=}T_{blu}|Y_{blu}) =$
$P(Y_{blu}|X_{blu})P(X_{blu}) / P(Y_{blu})$



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

$P(X{=}T_{blu}|Y_{blu}) = P(Y_{blu}|X_{blu})P(X_{blu}) / P(Y_{blu}) = 0.8*0.9 / 0.74 = 0.97$

Affidabilità della stima - SW in Matlab

N = 10 ➔ $P(X{=}T_{blu}|Y_{blu}) = P(Y_{blu}|X_{blu})P(X_{blu}) / P(Y_{blu}) = 0.5$

N = 100 ➔ $P(X{=}T_{blu}|Y_{blu}) = P(Y_{blu}|X_{blu})P(X_{blu}) / P(Y_{blu}) = 0.3235$

N = 1,000 ➔ $P(X{=}T_{blu}|Y_{blu}) = P(Y_{blu}|X_{blu})P(X_{blu}) / P(Y_{blu}) = 0.4$

N = 10,000 ➔ $P(X{=}T_{blu}|Y_{blu}) = P(Y_{blu}|X_{blu})P(X_{blu}) / P(Y_{blu}) = 0.4157$

N = 100,000 ➔ $P(X{=}T_{blu}|Y_{blu}) = P(Y_{blu}|X_{blu})P(X_{blu}) / P(Y_{blu}) = 0.4173$

N = 1,000,000 ➔ $P(X{=}T_{blu}|Y_{blu}) = P(Y_{blu}|X_{blu})P(X_{blu}) / P(Y_{blu}) = 0.4158$

Possiamo dare degli intervalli di confidenza?
Quanto deve essere grande N per ottenere una certa confidenza?

Esempio - II

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 sensitività della mammografia è intorno al 90%:

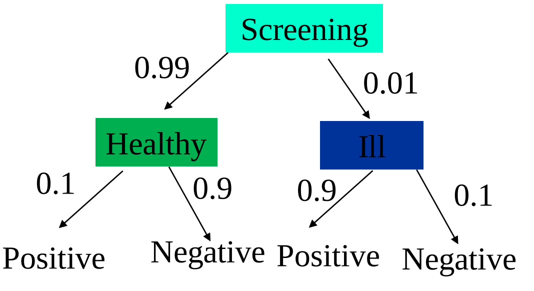
$$\text{sensitività} = \frac{n_{positive}}{N_{ill}} \quad \Rightarrow P(Y=positive \mid X=ill)$$

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

$$\text{specificità} = \frac{n_{negative}}{N_{healthy}} \quad \Rightarrow P(Y=negative \mid X=healthy)$$

Esempio II

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



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

P(Y=Positive) = P(Y=Positive | X=Ill) + P(Y=Positive | X=Healthy) =
0.009 + 0.099 = 0.108

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 - II

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

Applichiamo Bayes:
P(X=Ill | Y=Positive) = P(Y=Positive | X=Ill)P(X=Ill) / P(Y=Positive)

P(X = Ill) = 0.01

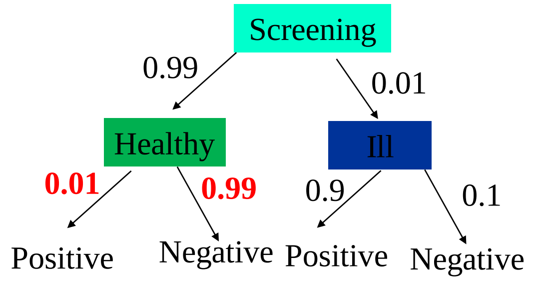
Il PPV (Positive Predictive Value) è:
P(X=Ill | Y=Positive) = P(Y=Positive | X=Ill)P(X=Ill) / P(Y=Positive) =
0.09 / 0.108 = 0.083 (8.3%)

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?

Esempio II

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



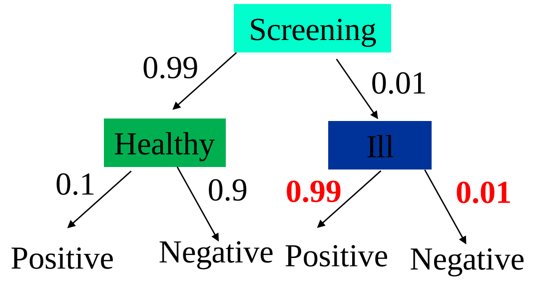
P(Y=Positive | X=Ill) = 0.9 * 0.01 = 0.009

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%

Esempio II

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



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%.

Riepilogo

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

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.

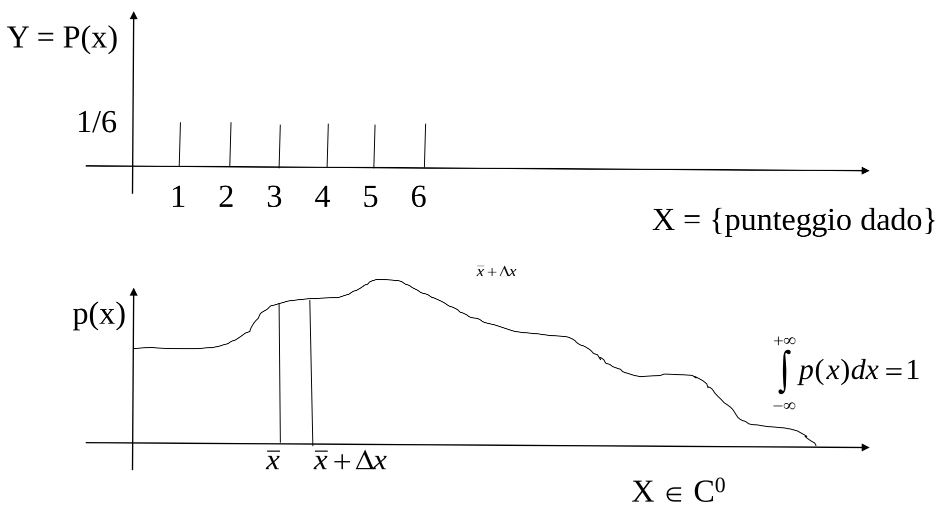
Viene utilizzata nei problemi inversi.

Overview

Esempi sul teorema di Bayes

Densità di probabilità

La probabilità nel caso continuo



$$\int_{-\infty}^{+\infty} p(x)dx = 1$$

$$P(x \in [\bar{x}, \bar{x}+\Delta x]) \int_{\bar{x}}^{\bar{x}+\Delta x} p(x)dx$$

Definizione di p(x)

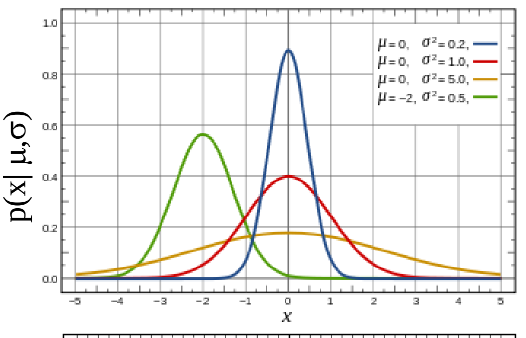
Caso discreto: prescrizione della probabilità per ognuno dei finiti valori che la variabile X può assumere: P(X).

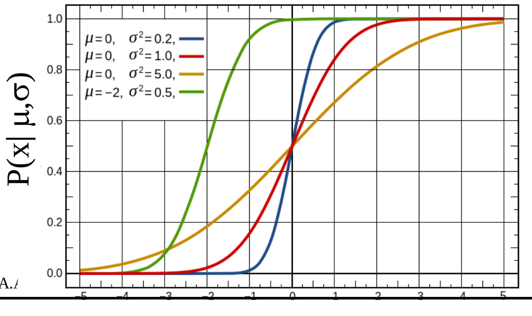
Caso continuo: i valori che X può assumere sono infiniti. Devo trovare un modo per definirne la probabilità. Descrizione **analitica** mediante la funzione densità di probabilità. Si considera la probabilità che x cada in un certo intervallo.

Valgono le stesse relazioni del caso discreto, dove alla somma si sostituisce l'integrale.

$$P(X = x \in [\bar{x}, \bar{x}+\Delta x]) \int_{\bar{x}}^{\bar{x}+\Delta x} \int_{-\infty}^{+\infty} p(x,y)dxdy$$

Distribuzioni notevoli: la Gaussiana





$$p(x \mid \mu, \sigma) = \frac{1}{\left(\sqrt{2\pi\Sigma^2}\right)^D} \cdot \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\Sigma}\right)^2\right]$$

D = dimensione, in questo caso D = 1

Pr(| X-μ | < σ) = 0.68268

Pr(| X-μ | < 2σ) = 0.95452

Pr(| X-μ | < 3σ) = 0.9973

I momenti di una variabile statistica

$\mu^k(X) = \int_{-\infty}^{+\infty} (x-a)^k \ p(x)dx$ Momento rispetto ad a, solitamente alla media

$E[X] = \int_{-\infty}^{+\infty} x \ p(x)$ Valore atteso (Expeted value) di X = media distribuzione

$E\left[(X-\mu)^2\right] = \int_{-\infty}^{+\infty} (x-\mu)^2 \ p(x)$ Varianza (σ^2)

$E\left[(X-\mu)^3\right] = \int_{-\infty}^{+\infty} (x-\mu)^3 \ p(x)$ Asimmetria

$E\left[(X-\mu)^4\right] = \int_{-\infty}^{+\infty} (x-\mu)^4 \ p(x)$ Kurtosi – peso delle code di p(x)

Overview

Esempi sul teorema di Bayes

Densità di probabilità