Classificazione Bayesiana

Esempio. Specie di piante : {Asteracee, Orchidacee, Rubiacee} = $\mathcal{Y}$ (classi "cause")

Colori dei fiori : {viola, bianco, giallo} = $\mathcal{X}$ (valori osservabili)

100 osservazioni :

- 40 sono Asteracee, di cui 5 con fiori viola, 20 con fiori bianchi, 15 con fiori gialli
- 40 sono Orchidacee, di cui 20 " " viola, 10 " " bianchi, 10 " " gialli
- 20 sono Rubiacee, di cui tutte con fiori gialli

Un classificatore è una funzione $h : \mathcal{X} \to \mathcal{Y}$

Data un'osservazione $x \in \mathcal{X}$, vorremmo che $h(x)$ massimizzasse la probabilità di corretta classificazione : $\Pr[h(X) = Y \mid X = x]$

(X, Y)

Il classificatore bayesiano associa ad ogni $x \in \mathcal{X}$ la classe $\hat{y} \in \mathcal{Y}$ che massimizza tale probabilità :

$$\boxed{\hat{y} = \underset{y_j}{\operatorname{argmax}} \Pr_{Y}[\hat{y} = Y \mid x]}$$

Codici di canale :

$\underset{\sim}{X} \longrightarrow$ canale $\longrightarrow Y \longrightarrow \boxed{\text{decodif.}}$

ingresso del canale

y

Criterio $\underbrace{\text{della decodifica a massima verosim.}}$:

data un'uscita del canale $y \in \mathcal{Y}$,

restituisci la parola di codice $\hat{x}$

che massimizza $p(\hat{x} | y)$:

$$\hat{x} = \arg\max_{x_i} \; p(x_i | y)$$

fisso (dato)

Classificazione

$\underset{\sim}{Y} \longrightarrow$ natura $\longrightarrow X \longrightarrow \boxed{\text{classificatore}}$

classe (etichetta) (↔ specie di piante)

colore del fiore

Classific. bayesiana,

data l'osservazione $x \in \mathcal{X}$,

restituisci la classe (etichetta) $\hat{y} \in \mathcal{Y}$

che massimizza $p(\hat{y} | x)$:

$$\hat{y} = \arg\max_{y_j} \; p(y_j | x) .$$

$p(x,y)$	viola	bianchi	gialli
Asteracee	5%	20%	15%
Orchidacee	20%	10%	10%
Rubiacee	0%	0%	20%

(1/4 del totale)

$p(y\|x)$	viola	bianchi	gialli
Asteracee	1/5	2/3	3/9
Orchidacee	4/5	1/3	2/9
Rubiacee	0	0	4/9
	1/5	1/3	5/9

$$p(y|x) = \frac{p(x,y)}{p(x)}$$

$$\frac{5}{25}, \quad \frac{20}{25}, \quad \frac{0}{25}$$

Quindi $h(\text{viola}) = $ Orchidacee è l'ipotesi bayesiana per l'osservazione $x = $ viola

$h(\text{bianco}) = $ Asteracee " " " " $x = $ bianco

$h(\text{giallo}) = $ Rubiacee " " " " $x = $ giallo

Qual è la prob. di errore complessiva?

$p(X = viola) = 25/100 = 1/4$

$p(X = bianco) = 30/100 = 3/10$

$p(X = giallo) = 45/100 = 9/20$

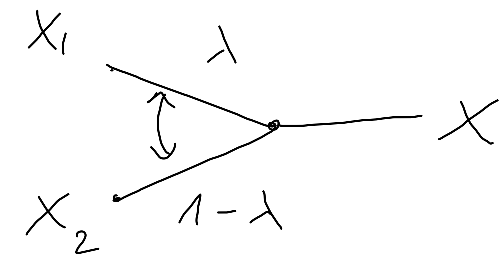
Prob. di errata classificazione $= 1/4 \cdot 1/5 + 3/10 \cdot 1/3 + 9/20 \cdot 5/9 = 8/20 = 40\%$.

① Consideriamo due v.a. X_1 e X_2.

Definiamo una terza v.a. $X = \begin{cases} X_1 & \text{con prob. } \lambda \\ X_2 & \text{con prob. } 1-\lambda \end{cases}$ (λ parametro fissato) $\lambda \in [0,1]$

Dimostrare che $H(X) \geq \lambda H(X_1) + (1-\lambda) H(X_2)$.

Definisco una v.a. Z binaria:



$$Z = \begin{cases} 1 & \text{se } X = X_1 \quad (\text{prob. } \lambda) \\ 2 & \text{se } X = X_2 \quad (\text{prob. } 1-\lambda) \end{cases}$$

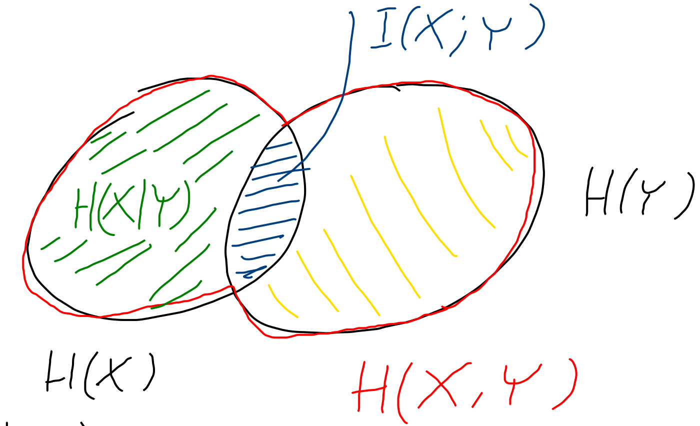
$$H(X) \geq H(X|Z) = \underbrace{Pr[Z=1]}_{\lambda} \underbrace{H(X|Z=1)}_{H(X_1)} + \underbrace{Pr[Z=2]}_{1-\lambda} \underbrace{H(X|Z=2)}_{H(X_2)}.$$

aggiungere un condizionamento non può aumentare l'entropia di X

proprietà entropia condizionata

② Consideriamo due v.a. X e Y.

Sia $\rho = \dfrac{I(X;Y)}{H(X,Y)}$



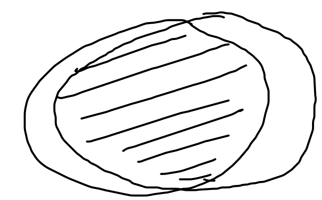
Dimostrare che:

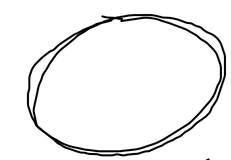
$H(X,Y) = H(X|Y) + I(X;Y) + H(Y|X)$

(a) $0 \le \rho \le 1$

(b) Dare una condiz. necess. e suff. affinché $\rho = 0$

(c) " " " " " " " $\rho = 1$





$H(X) = H(Y) = H(X,Y) = I(X;Y)$

(a) $H(X,Y) > 0$ e $I(X;Y) \ge 0 \implies \rho \ge 0$

(* a meno che $X = \text{cost.}, Y = \text{cost.}$)

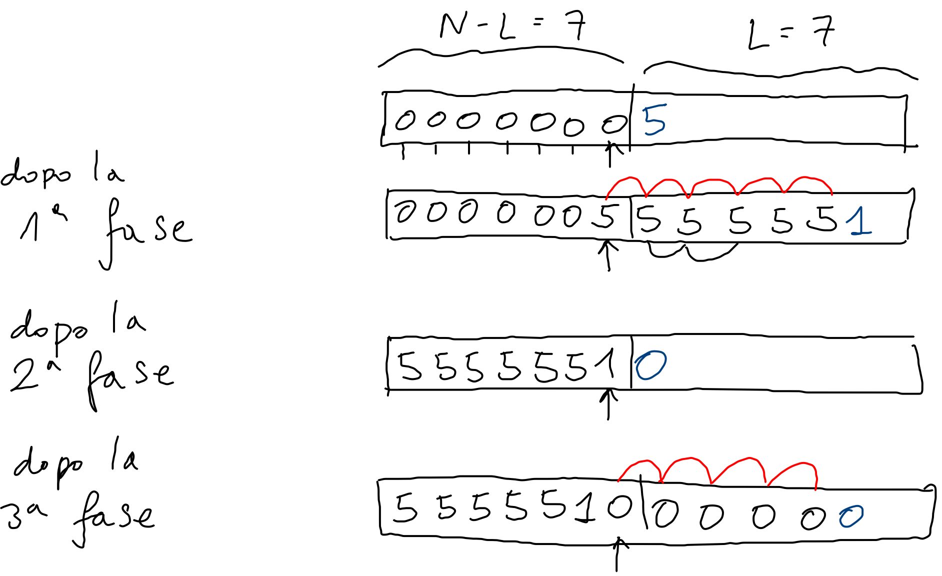
$I(X;Y) \le H(X,Y) \implies \rho \le 1$

(b) $\rho = 0 \iff I(X;Y) = 0 \iff X$ e Y sono v.a. indipendenti

→ (c) $\rho = 1 \iff I(X;Y) = H(X,Y) \iff H(X|Y) = 0$ e $H(Y|X) = 0 \iff$ La X è una funz. di Y, La Y è una funz. di X, X e Y sono in corrispondenza biunivoca.

③ Decodifica di Ziv-Lempel (LZ77) della stringa

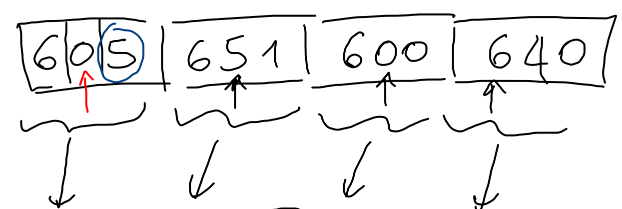
605 651 600 640 in alfabeto decimale ($N = 14$, $L = 7$, $K = 10$)



<u>Lunghezza dei blocchi:</u>

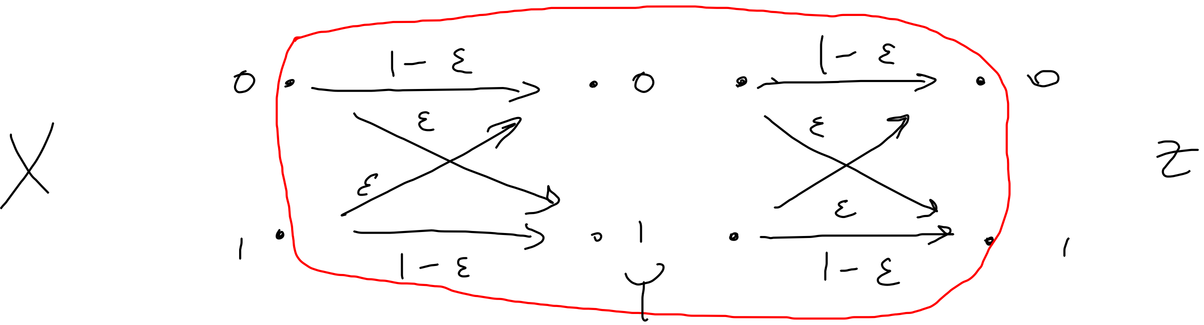
- Parte puntatore: $\lceil \log_{10}(N-L) \rceil = \lceil \log_{10} 7 \rceil = \underline{1}$
- Parte lunghezza: $\lceil \log_{10} L \rceil = \lceil \log_{10} 7 \rceil = \underline{1}$
- 1 simbolo aggiuntivo: $\underline{1}$

Totale: 3



Sequenza decodificata è 5 | 555551 | 0 | 00000

④ Concatenazione di canali BSC(ε)



Scrivere la matrice di transizione del canale ottenuto concatenando due canali BSC(ε).

Matrice del primo canale : $\Gamma = \begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix}$

Matrice del secondo canale : $\Gamma =$ "

Matrice di transizione del canale concatenato

$p_Y = p_X \Gamma$ $\qquad p_Z = p_Y \cdot \Gamma \implies p_Z = (p_X \Gamma) \cdot \Gamma = p_X \Gamma^2$

vettore riga

$$\Gamma^2 = \begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix}\begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix} = \begin{pmatrix} (1-\varepsilon)^2+\varepsilon^2 & 2\varepsilon(1-\varepsilon) \\ 2\varepsilon(1-\varepsilon) & (1-\varepsilon)^2+\varepsilon^2 \end{pmatrix} \cdot \square$$

$$= \begin{pmatrix} \frac{1}{2}\left(1+(1-2\varepsilon)^2\right) & \frac{1}{2}\left(1-(1-2\varepsilon)^2\right) \\ \frac{1}{2}\left(1-(1-2\varepsilon)^2\right) & \frac{1}{2}\left(1+(1-2\varepsilon)^2\right) \end{pmatrix}$$

In generale, per induzione su n,

$$\Gamma^n = \begin{pmatrix} \frac{1}{2}\left(1+(1-2\varepsilon)^n\right) & \frac{1}{2}\left(1-(1-2\varepsilon)^n\right) \\ \frac{1}{2}\left(1-(1-2\varepsilon)^n\right) & \frac{1}{2}\left(1+(1-2\varepsilon)^n\right) \end{pmatrix}$$

$$\xrightarrow[(\text{se } \varepsilon \in (0, 1/2))]{n \to \infty} \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

Canale inutile.