Entropia e Teoria dell'Informazione

Claudio Meggio

Anno Accademico
2016/2017

Università degli Studi di
Trento

Indice

Introduzione

Esempio

i. Quando vado in palestra mi alleno
ii. Il vincitore delle prossime elezioni sarà Claudio Baglioni
iii. QUER W LKS E W

Esempio

i. Quando vado in palestra mi alleno
ii. Il vincitore delle prossime elezioni sarà Claudio Baglioni
iii. QUER W LKS E W

Definizione Informazione

In uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ definiamo la funzione **informazione** $I : \mathcal{F} \to \mathbb{R}^+$ come:

$$I(E) = -\log_a(\mathbb{P}(E)).$$

Definizione:
U viene detta **misura di incertezza** se soddisfa le seguenti:

- $U(X)$ è un massimo quando ha distribuzione uniforme
- $U(p_1 ... p_n, 0) = U(p_1 ... p_n)$
- $U(p_1 .... p_n)$ è continua per tutti i suoi argomenti.
- Presa Y variabile casuale allora
  $U(X, Y) = U_X(Y) + U(X)$
  dove $U_X(Y) = \sum_{j=1}^{n} p_j U(Y|X = j)$

Definizione Entropia

Data X variabile casuale definiamo la sua **Entropia** come:

$$H(X) := \mathbb{E}(I(X)) = -\sum_{j=1}^{n} p_j \log(p_j)$$

Teorema:

$U(X)$ è una misura di incertezza se e solo se

$$U(X) = KH(X), \quad K > 0$$

Proprietà

Teorema

$$H(x) \leq \log(n)$$

Con l'uguaglianza sse X ha distribuzione uniforme

Dimostrazione:

$$\begin{aligned} H(x) - \log(n) &= -\frac{1}{ln(2)}\left(\sum_{j=1}^{n} p_j ln(p_j) + ln(n)\right) \\ &= -\frac{1}{ln(2)}\left(\sum_{j=1}^{n} p_j (ln(p_j) + ln(n))\right) \\ &\leq \frac{1}{ln(2)}\left(\sum_{j=1}^{n} p_j \left(\frac{1}{p_j n} - 1\right)\right) \leq 0 \end{aligned}$$

Definizione Entropia Condizionata

$$H_X(Y) := \mathbb{E}[H.(Y)] = \sum_{j=1}^{n} p_j H_j(Y)$$

dove $H_j(Y) := -\sum_{k=1}^{m} p_j(k) \log(p_j(k))$

Entropia Condizionata

Definizione Entropia Condizionata

$$H_X(Y) := \mathbb{E}[H.(Y)] = \sum_{j=1}^{n} p_j H_j(Y)$$

dove $H_j(Y) := -\sum_{k=1}^{m} p_j(k) \log(p_j(k))$

Disuguaglianza di Shannon

$$H_X(Y) \leq H(Y)$$

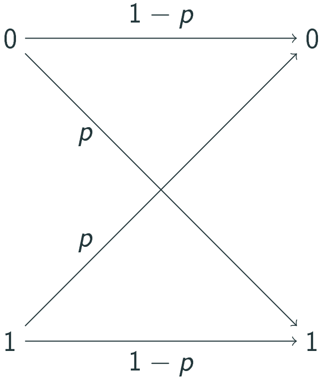
Dimostrazione:

Disugauglianza di Jensen con

$$\lambda_j = p_j \quad f(x) = x \log(x) \quad x_j = p_j(k)$$

Canale Binario simmetrico

$$SORGENTE \rightarrow CANALE \rightarrow RICEVENTE$$



Definizione Capacità

Viene definita capacità di un canale la quantità

$$\begin{aligned} C := & \max_{\{p_1 \dots p_n\}} I(S,R) \\ = & \max_{\{p_1 \dots p_n\}} (H(R) - H_S(R)) \end{aligned}$$

Capacità canale simmetrico binario $C = 1 - H_S(R)$

Velocità

Velocità di trasmissione è definita come il numero di bits d'informazione che vengono trasmessi attraverso il canale. Nel nostro caso (un simbolo al secondo) la velocità è data da $I(R, S)$ commento sulla velocità in rapporto alla lunghezza delle parole di un codice

Errore

Probabilità media d'errore

$$P(E) = \sum_{i=1}^{N} P_{x_j}(E) p_j$$

Distanza di Hamming

Numero di simboli che differiscono nelle due Regola di decisione

da errore se ci sono più parole nella stessa sfera, oppure non ve ne sono

Lemmi preparativi

Lemma 1

Per ogni fissato $\delta_1 > 0$, scelto d sufficientemente grande vale:

$$\mathbb{P}(A) \leq \delta_1$$

Lemma 2

Siano ρ e δ_2 due numeri reali non negativi e supponiamo che le parole del codice siano $M = 2^{d(C-\rho)}$ dove $C = 1 - H_b(p)$ è la capacità del canale allora, per d sufficientemente grande vale:

$$\mathbb{P}(B) \leq \delta_2$$

Teorema di Shannon

Dati $\delta, \rho > 0$ possiamo trovare un codice tale per cui se la velocità di trasmissione in un canale binario simmetrico è $V = C - \rho$ allora

$$\mathbb{P}(E) < \delta$$

commenti

Conclusioni

FINE

GRAZIE DELL'ATTENZIONE