TEORIA DELL'INFORMAZIONE E INFERENZA STATISTICA

Sia $\{p_\vartheta(x)\}_\vartheta$ una famiglia di d.p. parametrizzate dal parametro $\vartheta \in \mathbb{R}$

Sia X una v.a. campionata da p_ϑ $(X \sim p_\vartheta)$

Una statistica è una funzione $T(X)$ del campione X

Esempio: Media campione: $T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$

Poiché $T(X)$ è una funzione del campione, vale la catena di Markov:

$$\vartheta \to X \to T(X) \quad (\text{poiché } Pr[T(X) | \vartheta, X] = Pr[T(X) | X])$$

Per il 2° teorema di elaborazione dati, $I(\vartheta; T(X)) \leq I(\vartheta; X)$.

Quando $I(\vartheta; T(X)) = I(\vartheta; X)$, la statistica $T(X)$ si dice <u>statistica sufficiente</u>.

(per ricostruire ϑ, posso ignorare X e tenere solo $T(X)$)

→ è vero quando $I(\vartheta; T(X)) \geq I(\vartheta; X)$, cioè quando $Pr[X | \vartheta, T(X)] = Pr[X | T(X)]$

ovvero quando $\vartheta \to T(X) \to X$

Esempio. $X_1, \dots, X_n$ v.a. $\in \{0,1\}$ indipendenti con $p(1) = \vartheta$, $p(0) = 1-\vartheta$ $(\vartheta \in [0,1])$

$$Pr[(X_1,\dots,X_n) = (x_1,\dots,x_n)] = \vartheta^{\overbrace{\sum_i x_i}^{\text{Numero di "1"}}} \cdot (1-\vartheta)^{\overbrace{n-\sum_i x_i}^{\text{Numero di "0"}}}$$

$(x_1,\dots,x_n)$: quanti 1 e quanti 0 ci sono?

$$Pr[10001] = \vartheta^2(1-\vartheta)^3$$

$$Pr[01001] = \vartheta^2(1-\vartheta)^3$$

$$T(X_1,\dots,X_n) = \sum_{i=1}^n X_i$$

$$Pr\left[(X_1,\dots,X_n) = (x_1,\dots,x_n) \,\middle|\, \overbrace{\sum_{i=1}^n X_i}^{T(X_1,\dots,X_n)} = k\right] = \begin{cases} 1/\binom{n}{k} & \text{se } \sum_i x_i = k \\ 0 & \text{se } \sum_i x_i \neq k \end{cases}$$

non dipende da ϑ!

$\Rightarrow T(X_1,\dots,X_n) = \sum_i X_i$ è una statistica sufficiente

STIMA A MASSIMA VEROSIMIGLIANZA

Ho una famiglia $\{p_\vartheta(x)\}_\vartheta$ di d.p. ; $p_{\vartheta^*}(x)$ è il modello corretto ma non conosco ϑ^*

Quale scelta di ϑ è maggiormente consistente con le osservazioni?

$$\to D(p_{\vartheta^*} \| p_\vartheta) = \underset{X\sim p_{\vartheta^*}}{E}\left[\log \frac{p_{\vartheta^*}(X)}{p_\vartheta(X)}\right] = \overbrace{\underset{X\sim p_{\vartheta^*}}{E}\left[\log p_{\vartheta^*}(X)\right]}^{\text{dipende solo da } \vartheta^*} - \underset{X\sim p_{\vartheta^*}}{E}\left[\log p_\vartheta(X)\right]$$

$$= \text{costante} - \underset{X\sim p_{\vartheta^*}}{E}\left[\log p_\vartheta(X)\right]$$

(costante rispetto a ϑ)

Non lo conosco ma posso stimarlo: per la legge dei grandi numeri, $\frac{1}{n}\sum_{i=1}^n \log p_\vartheta(X_i) \hookrightarrow \underset{X\sim p_{\vartheta^*}}{E}\left[\log p_\vartheta(X)\right]$

$$\hat{D}(p_{\vartheta^*} \| p_\vartheta) = \text{costante} - \frac{1}{n}\sum_{i=1}^n \log p_\vartheta(X_i)$$

Scelgo ϑ per minimizzare la divergenza $\hat{D}$

$$\min_{\vartheta} \hat{D}(p_{\vartheta^*} \| p_\vartheta) \iff \min_{\vartheta} \left[-\frac{1}{n} \sum_{i=1}^{n} \log p_\vartheta(X_i) \right]$$

$$\iff \max_{\vartheta} \left[\sum_{i=1}^{n} \log p_\vartheta(X_i) \right]$$

$$\iff \max_{\vartheta} \underbrace{\prod_{i=1}^{n} p_\vartheta(X_i)}_{\text{funzione di verosimiglianza (likelihood) } \mathcal{L}(\vartheta | X)}$$

log è una funz monotone

→ Principio di massima verosimiglianza

Esempio: $\mathcal{L}(\vartheta | \vec{X}) = \prod_{i=1}^{n} p_\vartheta(X_i) = \vartheta^k (1-\vartheta)^{n-k}$ dove $k = \sum_{i=1}^{n} X_i$

$p_\vartheta(X_i)$: ϑ (se $X_i = 1$), $(1-\vartheta)$ (se $X_i = 0$)

$k - k\vartheta = n\vartheta - k\vartheta \iff \vartheta = k/n$

$= 0 \iff k(1-\vartheta) = (n-k)\vartheta$

$$\mathcal{L}' = k\vartheta^{k-1}(1-\vartheta)^{n-k} + (n-k)\vartheta^k (1-\vartheta)^{n-k-1}(-1) = \vartheta^{k-1}(1-\vartheta)^{n-k-1} \underbrace{[k(1-\vartheta) - (n-k)\vartheta]}_{\neq 0 \quad \neq 0}$$

$\mathcal{L}' = 0$

Esercizio. Codifica run-length

Siano $X_1, \dots, X_n$ v.a. $\in \{0,1\}$ ma non necessariamente dipendenti.

Si considerino le lunghezze delle run $\vec{R} = (R_1, R_2, \dots, \dots)$

Esempio: $\vec{X} = \underbrace{00}_{3}\underbrace{11}_{2}\underbrace{00}_{2}\underbrace{1}_{1}\underbrace{00}_{2}$, $\vec{R} = (3,2,2,1,2)$? $\to 000110010$ (6) ? $\to 111001101$ (11)

Si confrontino $H(X_1, \dots, X_n)$ (cioè $H(\vec{X})$), $H(\vec{R})$, e $H(X_n, \vec{R})$,

e limitare le differenze tra queste 3 quantità.

Svolg. $\vec{R}$ è funzione di $\vec{X}$, quindi $H(\vec{R}) \le H(\vec{X}) = H(X_1, \dots, X_n)$.

Inoltre: $(X_n, \vec{R})$ sono funz. di $\vec{X}$ $\Rightarrow H(X_n, \vec{R}) \le H(\vec{X})$

$\vec{X}$ è funz. di $(X_n, \vec{R})$ $\Rightarrow H(\vec{X}) \le H(X_n, \vec{R})$ $\Big\}$ $H(\vec{X}) = H(X_n, \vec{R})$

$$H(\vec{R}) \le \underbrace{H(\vec{X}) = H(X_n, \vec{R})}_{} = H(\vec{R}) + H(X_n|\vec{R}) \le H(\vec{R}) + \overbrace{H(X_n)}^{\le 1} \le H(\vec{R}) + 1$$

Esercizio. Siano $X_1, X_2, \dots, X_n$ di v.a. indipendenti con la stessa distribuzione, con entropia H.

Sia $C_n(t) = \{ \vec{x} \in \mathcal{X}^n : p(\vec{x}) \geq 2^{-nt} \}$

(a) Mostrare che $|C_n(t)| \leq 2^{nt}$

(b) Per quali valori di t si ha $\Pr[(X_1, \dots, X_n) \in C_n(t)] \longrightarrow 1$?

(a) $\Pr[(X_1, \dots, X_n) \in C_n(t)] \leq 1$ quindi

$$\boxed{1} \geq \Pr[(X_1, \dots, X_n) \in C_n(t)] \geq |C_n(t)| \min_{\vec{x} \in C_n(t)} p(\vec{x}) \geq |C_n(t)| \cdot 2^{-nt}$$

$\log p \rightsquigarrow -nH$
$p \rightsquigarrow 2^{-nH}$

$\Rightarrow C_n(t) \leq 2^{nt}$.

(b) Per il principio di equipartizione asintotica, $-\frac{1}{n} \log p(\vec{X}) \rightsquigarrow H$ (autoinformaz. normalizzate)

ovvero $p(\vec{X})$ è vicina a 2^{-nH} con prob. che tende a 1.

Quindi $\Pr[p(\vec{X}) \geq 2^{-nt}] \longrightarrow 1$ se $H < t$, $\longrightarrow 0$ se $H > t$.

Esercizio. Codifica di sorgente

Una sorgente senza memoria e stazionaria emette una sequenza di cifre binarie

con probabilità: $p(1) = 0.005$
$p(0) = 0.995$

0100110000010...
100

Prendiamo le cifre 100 alla volta (blocchi di dimensione 100)

e se il blocco ha al più 3 cifre 1, lo codifichiamo con una parola di codice binaria.

(altrimenti lo ignoriamo).

(a) Se tutte le parole di codice hanno uguale lunghezza, qual è la loro minima lunghezza possibile?

(b) Calcolare la prob. di osservare un blocco a cui non è stata associata una parola di codice.

(a) Le seq. di 100 bit con al più 3 cifre 1?

$$\binom{100}{0} + \binom{100}{1} + \binom{100}{2} + \binom{100}{3} = 1 + 100 + \frac{100 \cdot 99}{2} + \frac{100 \cdot 99 \cdot 98}{3 \cdot 2} = 166751$$

Lunghezza minima: $\lceil \log_2 166751 \rceil = 18$. $(H(X) \simeq 4.54)$

(b) La prob. di avere al più 3 cifre 1 in un blocco è:

$$\sum_{i=0}^{3} \binom{100}{i} (0.005)^i (0.995)^{100-i} \cong 0.99\,833\ldots$$

La prob. di non poter codificare è $1 - 0.99833\ldots < 1\%$.