Teoria dell'Informazione

Simone Alessandro Casciaro
15 Ottobre 2024

Lezione 5: 1° Teorema di Shannon

Upper Bound del Valore Atteso

Ipotesi

Data una sorgente $< \mathbb{X}, \mathbb{P} >$ e la variabile aleatoria $X : \mathbb{X} \to \mathbb{R}$
Dato c codice di Shannon (quindi istantaneo) con lunghezza delle codifiche $l_i = l_c(x_i) \quad \forall i = 1, \ldots, m$ tali che $l_i = \lceil \log_d \frac{1}{p_i} \rceil$

Tesi

$$\mathbb{E}(l_c) < H_d(X) + 1$$

Dimostrazione

$\mathbb{E}(l_c) = \sum_{i=1}^{m} p_i l_i = \sum_{i=1}^{m} pi \lceil \log_d \frac{1}{p_i} \rceil < \sum_{i=1}^{m} p_i \left(\log_d \frac{1}{p_i} + 1 \right) = \sum_{i=1}^{m} p_i \log_d \frac{1}{p_i} + \sum_{i=1}^{m} p_i = H_d(X) + 1$

Problema!
Abbiamo dimostrato che l'errore che commettiamo stimando il valore atteso con l'entropia è compreso tra 0 e 1 (Dato che $H_d(X) \leq \mathbb{E}(l_c) < H_d(X) + 1$). Tuttavia, questo risultato è relativo alla codifica di ogni singolo simbolo, dunque il problema si moltiplica per il numero di simboli nel messaggio.

$$l_c(x_1, \ldots, x_m) = \sum_{i=1}^{m} l_c(x_i) = \sum_{i=1}^{m} \left\lceil \log_d \frac{1}{p_i} \right\rceil$$

> Esempio
>
> | x_1 | x_2 | x_3 |
> |---|---|---|
> | 2.5 | 2.1 | 2.1 |
> | $\lceil 2.5 \rceil +$ | $\lceil 2.1 \rceil +$ | $\lceil 2.1 \rceil = 9$ |
>
> Usiamo un trucco. Invece di sommare le approssimazioni, prima sommiamo e poi approssimiamo.
> $\lceil 2.5 + 2.1 + 2.1 \rceil = \lceil 6.7 \rceil = 7$ e quindi risparmiamo 2 bit.

Costruiamo dunque un nuovo codice

$$C_n : \mathbb{X}^n \to D^+$$

che codifica blocchi di simboli della sorgente di lunghezza fissa n.

Idealmente questo codice è migliore di quelli visti precedentemente, ma la complessità computazionale del codice aumenta con l'aumentare di n.

Entropia Congiunta

Supponiamo di avere una sorgente $< \mathbb{X}, \mathbb{P} >$, dalla quale estraiamo una tupla $(x_1, \dots, x_n)$ che vogliamo codificare usando la funzione $C_n : \mathbb{X}^n \to D^+$.
Si ricorda che Shannon, per semplificare i calcoli, assume che i simboli di un messaggio siano indipendenti tra loro e, dunque, che

$$p(x_1, \dots, x_n) = \prod_{i=1}^{n} p(x_i) = \mathbb{P}_n(x_1, \dots, x_n)$$

Definiamo una nuova sorgente $< \mathbb{X}^n, \mathbb{P}_n >$ e ne calcoliamo l'entropia. Per farlo, supponiamo di avere n variabili aleatorie $X_1, \dots, X_n$, tutte indipendenti e identicamente distribuite, che estraggono dalla stessa sorgente.

> Premessa utile per il calcolo successivo
>
> $$\log_d \frac{1}{\mathbb{P}(x_1, \dots, x_n)} = \log_d \frac{1}{\prod_{i=1}^{n} p(x_i)} = \sum_{i=1}^{n} -(\log_d p_i) = \sum_{i=1}^{n} \log_d \frac{1}{p_i}$$

$$
\begin{aligned}
H_d(X_1, \ldots, X_n) = \sum_{X_1, \ldots, X_n} \mathbb{P}(x_1, \ldots, x_n) \log_d \frac{1}{\mathbb{P}(x_1, \ldots, x_n)} \\
= \sum_{X_1} \sum_{X_2} \cdots \sum_{X_n} \left(\prod_{i=1}^{n} p(x_i) \right) * \sum_{i=1}^{n} \log_d \frac{1}{p(x_i)}
\end{aligned}
$$

Questa formula risulta complicata se pensata al caso generale n. Per comprenderla, proviamo a vedere cosa succede nel caso $\boxed{n = 2}$

$$
\begin{aligned}
H_d(X_1, X_2) &= \sum_{X_1} \sum_{X_2} \prod_{i=1}^{2} p(x_i) \sum_{i=1}^{2} \log_d \frac{1}{p(x_i)} \\
&= \sum_{X_1} \sum_{X_2} \prod_{i=1}^{2} \left(\log_d \frac{1}{p(x_1)} + \log_d \frac{1}{p(x_2)} \right) \\
&= \sum_{X_1} \sum_{X_2} p(x_1) * p(x_2) * \left(\log_d \frac{1}{p(x_1)} + \log_d \frac{1}{p(x_2)} \right) \\
&= \sum_{X_1} \sum_{X_2} \left(p(x_1) * p(x_2) * \log_d \frac{1}{p(x_1)} + p(x_1) * p(x_2) \log_d \frac{1}{p(x_2)} \right) \\
&= \sum_{X_1} \sum_{X_2} p(x_1) * p(x_2) * \log_d \frac{1}{p(x_1)} + \sum_{X_1} \sum_{X_2} p(x_1) * p(x_2) * \log_d \frac{1}{p(x_2)} \\
&= \sum_{X_2} p(x_2) \sum_{X_1} p(x_1) \log_d \frac{1}{p(x_1)} + \sum_{X_1} p(x_1) \sum_{X_2} p(x_2) \log_d \frac{1}{p(x_2)} \\
&= \sum_{X_1} p(x_1) \log_d \frac{1}{p(x_1)} + \sum_{X_2} p(x_2) \log_d \frac{1}{p(x_2)} \\
&= H_d(X_1) + H_d(X_2)
\end{aligned}
$$

Dunque, in generale:

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

ma dato che le varie X_i seguono la stessa distribuzione di probabilità di X, vale che

$$H_d(X_1, \dots, X_n) = \sum_{i=1}^{n} H_d(X) = nH_d(X)$$

Primo Teorema di Shannon

Ipotesi

Sia $C_n : \mathbb{X}^n \rightarrow D^+$ la codifica di un codice a blocchi di Shannon $d-$ ario per la sorgente $< \mathbb{X}, \mathbb{P} >$ tale che $l_{C_n}(x_1, \dots, x_n) = \left\lceil \log_d \frac{1}{\mathbb{P}_n(x_1, \dots, x_n)} \right\rceil$

Tesi

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{E}(l_{C_n}) = H_d(X)$$

Dimostrazione

Da precedenti dimostrazioni, sappiamo che $H_d(X_1, \dots, X_n) \leq \mathbb{E}(l_{C_n}) < H_d(X_1, \dots, X_n) + 1$

Inoltre, sappiamo che $H_d(X_1, \dots, X_n) = nH_d(X)$

Unendo le due cose, si ha che $nH_d(X) \leq \mathbb{E}(l_{C_n}) < nH_d(X) + 1$

Per ottenere la tesi, si divide tutta la disequazione per n:

$$H_d(X) \leq \frac{1}{n} \mathbb{E}(l_{C_n}) < H_d(X) + \frac{1}{n}$$

Ponendo il $\lim_{n \to \infty}$, si ha che $\frac{1}{n} \mathbb{E}(l_{C_n})$ è compreso tra $H_d(X)$ e $H_d(X) +$ un infinitesimo, dunque

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{E}(l_{C_n}) = H_d(X)$$

Teorema: Upper Bound Valore Atteso di sorgente campionata

Ipotesi

Data una sorgente $< \mathbb{X}, \mathbb{P} >$ e $c : \mathbb{X} \to D^+$ un codice di Shannon tale che $l_c(x) = \left\lceil \log_d \frac{1}{q(x)} \right\rceil$, dove q è una distribuzione di probabilità stimata sulla sorgente $\mathbb{X}$ e associata alla variabile aleatoria $Y : \mathbb{X} \to \mathbb{R}$

Tesi

$$\mathbb{E}(l_c) < H_d(X) + 1 + D_d(X||Y)$$

Dimostrazione

$$\begin{aligned}
\mathbb{E}(l_c) &= \sum_{x \in X} p(x) \left\lceil \log_d \frac{1}{q(x)} \right\rceil \\
&< \sum_{x \in X} p(x) \left(\log_d \frac{1}{q(x)} + 1 \right) \\
&= \sum_{x \in X} p(x) \log_d \frac{1}{q(x)} + \sum_{x \in X} p(x) \\
&= \sum_{x \in X} p(x) \log_d \frac{1}{q(x)} + 1 \\
&= \sum_{x \in X} p(x) \log_d \left(\frac{1}{q(x)} \frac{p(x)}{p(x)} \right) + 1 \\
&= \sum_{x \in X} p(x) \log_d \frac{p(x)}{q(x)} + \sum_{x \in} p(x) \log_d \frac{1}{p(x)} + 1 \\
&= D_d(X||Y) + H_d(X) + 1
\end{aligned}$$

Esercizi

1. Ignorando la condizione di terminazione dell'algoritmo di Sardinas-Patterson dell'esercizio della scorsa lezione, provare ad andare avanti finché non si raggiunge un'ulteriore condizione di terminazione.
   $S_5 = \{ED\}$

$S_6 = \{D\}$
$S_7 = \{\}$

2. Costruire un codice istantaneo per una sorgente con le seguenti probabilità $\mathbb{P} = \left\{\frac{1}{3}, \frac{1}{4}, \frac{1}{6}, \frac{1}{8}, \frac{1}{9}, x\right\}$
$x = 1 - \frac{24 - 18 - 12 - 9 - 8}{72} = \frac{1}{72}$

Dunque le lunghezze sono:
$l_1 = 2$
$l_2 = 2$
$l_3 = 3$
$l_4 = 3$
$l_5 = 4$
$l_6 = 7$

E i relativi codici:
$c(x_1) = 00$
$c(x_2) = 01$
$c(x_3) = 100$
$c(x_4) = 101$
$c(x_5) = 1100$
$c(x_6) = 1101000$