Teoria dell'Informazione

Simone Alessandro Casciaro

11 Ottobre 2024

Lezione 4: Entropia

Definizione

Sia una sorgente $< \mathbb{X}, P >$ con $\mathbb{X} = \{x_1, \dots, x_m\}$ e $P = \{p_1, \dots, p_m\}$. Per comodità, indichiamo con p_i la probabilità con cui x_i compare nel messaggio.
Sia $X : \mathbb{X} \to \{a_1, \dots, a_m\} \subseteq \mathbb{R}$ una variabile aleatoria tale per cui $\mathbb{P}(X = a_i) = p_i$
Chiamiamo Entropia la funzione

$$H_d(X) = \sum_{i=1}^{m} p_i \log_d \frac{1}{p_i}$$

L'entropia dipende solo dalla distribuzione di probabilità di X e non da X stessa.

Cambio di base dell'entropia

Avendo a, b e p, ricordiamo che $\log_b p = \log_b a * \log_a p$
Quindi, per cambiare base all'entropia, possiamo:

$$\begin{aligned}
H_b(X) &= \sum_{i=1}^{m} p_i \log_b \frac{1}{p_i} \\
&= \sum_{i=1}^{m} p_i * \log_b a * \log_a \frac{1}{p_i} \\
&= \log_b a * \sum_{i=1}^{m} p_i \log_a \frac{1}{p_i} \\
&= \log_b a * H_a(X)
\end{aligned}$$

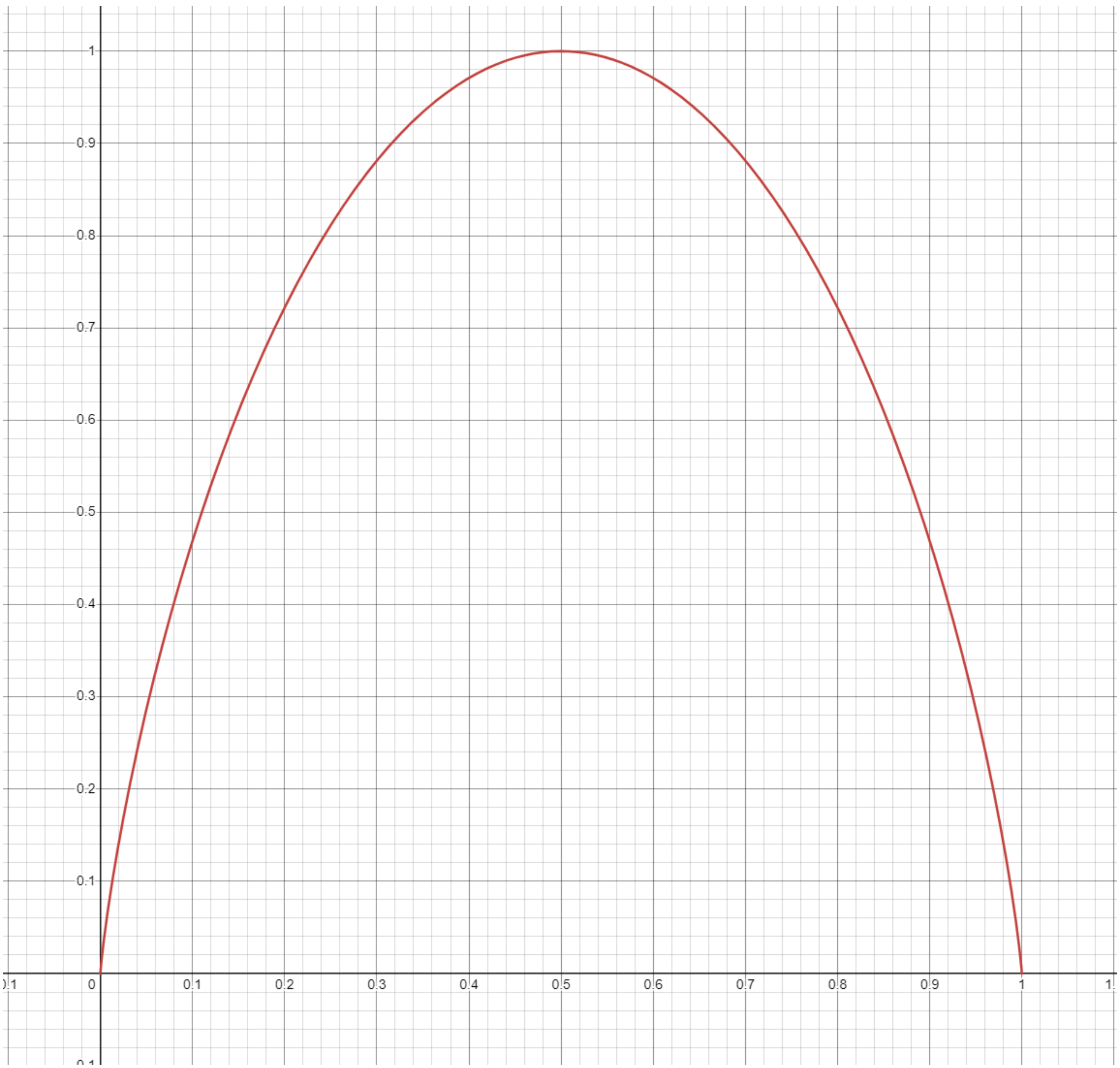
Dunque,

$$H_b(X) = \log_b a * H_a(X)$$

Rappresentazione sul Piano Cartesiano

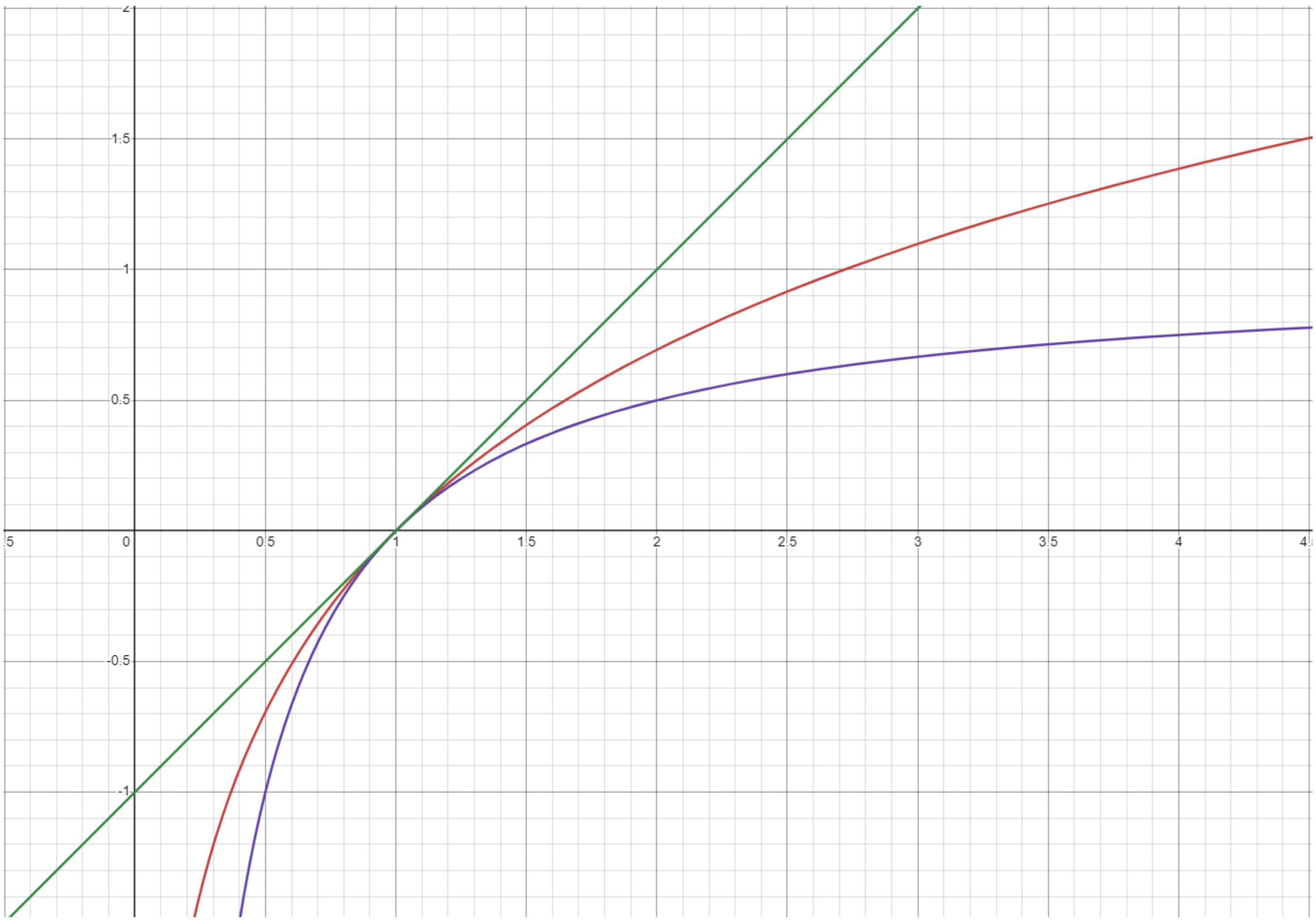
Sia X una variabile aleatoria Bernoulliana

$X : \mathbb{X} \to \{0, 1\}$ tale che $\mathbb{P}(X = 1) = p$ e $\mathbb{P}(X = 0) = 1 - p$



L'entropia risulta minima quando $p = 0$ oppure $p = 1$, mentre è massima quando $p = \frac{1}{2}$.
Questo indica che l'entropia è un indice che misura la quantità di informazione. Maggiore è l'entropia, maggiore è l'informazione spedita sul canale.

Maggiorante e minorante del $\ln x$



Come si vede dal disegno, abbiamo che $\forall x \in \mathbb{R} : 1 - \frac{1}{x} \leq \ln x \leq x - 1$
Questa informazione ci sarà utile per le dimostrazioni successive.

Enunciato: $H_d(X) \leq \log_d m \quad \forall d > 1$

Ipotesi

Sia X una variabile aleatoria

Tesi

$$H_d(X) \le \log_d m \quad \forall d > 1$$

$$H_d(X) = \log_d m \iff X \text{ ha una distribuzione uniforme}$$

Dimostrazione

Vogliamo dimostrare che $H_d(X) - \log_d m \le 0$

$$
\begin{aligned}
H_d(X) - \log_d m &= \sum_{i=1}^{m} p_i \log_d \frac{1}{p_i} - \log_d m && \text{Per definizione di Entropia} \\
&= \sum_{i=1}^{m} p_i \log_d \frac{1}{p_i} - \log_d m \sum_{i=1}^{m} p_i && \text{perché } \sum_{i=1}^{m} p_i = 1 \\
&= \sum_{i=1}^{m} p_i \log_d \frac{1}{p_i} - \sum_{i=1}^{m} p_i \log_d m \\
&= \sum_{i=1}^{m} p_i \left(\log_d \frac{1}{p_i} - \log_d m \right) \\
&= \sum_{i=1}^{m} p_i \left(\log_d \frac{1}{p_i m} \right) \\
&= \sum_{i=1}^{m} p_i \left(\log_d e * \ln \frac{1}{p_i m} \right) && \text{Cambio di base del logaritmo} \\
&= \sum_{i=1}^{m} p_i \left(\frac{1}{\ln d} * \ln \frac{1}{p_i m} \right) \\
&= \frac{1}{\ln d} \sum_{i=1}^{m} p_i \left(\ln \frac{1}{p_i m} \right) \\
&\leq \frac{1}{\ln d} \sum_{i=1}^{m} p_i \left(\frac{1}{p_i m} - 1 \right) && \ln x \leq x - 1 \quad \forall x \\
&= \frac{1}{\ln d} \sum_{i=1}^{m} \left(\frac{1}{m} - p_i \right) \\
&= \frac{1}{\ln d} \left(\sum_{i=1}^{m} \frac{1}{m} - \sum_{i=1}^{m} p_i \right) \\
&= \frac{1}{\ln d} (1 - 1) \\
&= 0
\end{aligned}
$$

Entropia Relativa

L'entropia relativa è una misura di distanza (non simmetrica!) tra due variabili aleatorie X e Y entrambe definite sul dominio S ma con due funzioni di probabilità diverse, che chiamiamo p_X e p_Y.

$$D_d(X||Y) = \sum_{s \in S} p_X(s) \log_d \frac{p_X(s)}{p_Y(s)}$$

NOTA BENE: $D_d(X||Y) \neq D_d(Y||X)$

Information Inequality

Ipotesi

X, Y due variabili aleatorie definite sul dominio S
$d > 1$

Tesi

$$D_d(X||Y) \geq 0$$

Dimostrazione

$$\begin{aligned}
D_d(X||Y) &= \sum_{s \in S} p_X(s) \log_d \frac{p_X(s)}{p_Y(s)} && \text{Per definizione di Entropia Relativa} \\
&= \sum_{s \in S} p_X(s) * \log_d e * \ln \frac{p_X(s)}{p_Y(s)} && \text{Cambio di base del logaritmo} \\
&= \log_d e \sum_{s \in S} p_X(s) \ln \frac{p_X(s)}{p_Y(s)} \\
&\geq \frac{1}{\ln d} \sum_{s \in S} p_X(s) \left(1 - \frac{p_Y(s)}{p_X(s)}\right) && \text{perch\'e } 1 - \frac{1}{x} \leq \ln x \quad \forall x \\
&= \frac{1}{\ln d} \sum_{s \in S} (p_X(s) - p_Y(s)) \\
&= \frac{1}{\ln d} \left(\sum_{s \in S} p_X(s) - \sum_{s \in S} p_Y(s) \right) \\
&= \frac{1}{\ln d}(1 - 1) \\
&= 0
\end{aligned}$$

Teorema $\mathbb{E}(l_c) \geq H_d(X)$

Ipotesi

$c : \mathbb{X} \to D^+$ codice istantaneo d-ario per una sorgente $< \mathbb{X}, \mathbb{P} >$

Tesi

$$\mathbb{E}(l_c) \geq H_d(X)$$

Dimostrazione

Sia $Y : \mathbb{X} \to \mathbb{R}$ una variabile aleatoria con funzione di probabilità $q(x) = \dfrac{d^{-l_c(x)}}{\sum\limits_{x' \in \mathbb{X}} d^{-l_c(x')}}$

Vogliamo dimostrare che $\mathbb{E}(l_c) - H_d(X) \geq 0$

$$\begin{aligned}
\mathbb{E}(l_c) - H_d(X) &= \sum_{x \in \mathbb{X}} p(x) l_c(x) - \sum_{x \in \mathbb{X}} p(x) \log_d \frac{1}{p(x)} \\
&= \sum_{x \in \mathbb{X}} p(x) \left(l_c(x) - \log_d \frac{1}{p(x)} \right) \\
&= \sum_{x \in \mathbb{X}} p(x) \left(\log_d d^{l_c(x)} - \log_d \frac{1}{p(x)} \right) \\
&= \sum_{x \in \mathbb{X}} p(x) \left(\log_d \frac{1}{d^{-l_c(x)}} + \log_d p(x) \right) \\
&= \sum_{x \in \mathbb{X}} p(x) \log_d \frac{p(x)}{d^{-l_c(x)}} \\
&= \sum_{x \in \mathbb{X}} p(x) \log_d \left(\frac{p(x)}{d^{-l_c(x)}} \frac{\sum\limits_{x' \in \mathbb{X}} d^{-l_c(x')}}{\sum\limits_{x' \in \mathbb{X}} d^{-l_c(x')}} \right) \\
&= \sum_{x \in \mathbb{X}} p(x) * \left[\log_d \left(p(x) \frac{\sum\limits_{x' \in \mathbb{X}} d^{-l_c(x')}}{d^{-l_c(x)}} \right) - \log_d \left(\sum_{x' \in \mathbb{X}} d^{-l_c(x')} \right) \right]
\end{aligned}$$

Dividiamo la sommatoria in due punti da studiare separatamente:

- 1

$$
\begin{aligned}
&\sum_{x \in \mathbb{X}} p(x) \log_d \left(p(x) \frac{\sum_{x' \in \mathbb{X}} d^{-l_c(x')}}{d^{-l_c(x)}} \right) \\
=&\sum_{x \in \mathbb{X}} p(x) \log_d \left(\frac{p(x)}{q(x)} \right) \\
=&D_d(X||Y) \\
\geq&\ 0
\end{aligned}
$$

- 2

$$
\begin{aligned}
&\sum_{x \in \mathbb{X}} p(x) \log_d \left(\sum_{x' \in \mathbb{X}} d^{-l_c(x')} \right) \\
&= \log_d \left(\sum_{x' \in \mathbb{X}} d^{-l_c(x')} \right) \sum_{x \in \mathbb{X}} p(x) \\
&= \log_d \left(\sum_{x' \in \mathbb{X}} d^{-l_c(x')} \right) \\
&\leq \log_d 1 \\
&= 0
\end{aligned}
$$

Di conseguenza, unendo i due punti, abbiamo qualcosa di positivo a cui viene sottratto qualcosa di negativo. $(1) - (2) \geq 0$

Algoritmo di Sardinas-Patterson

L'algoritmo di Sardinas-Patterson è il punto di riferimento per capire se un codice è univocamente decodificabile oppure no.
Si parte indicando con S_1 le parole del codice. Poi, si procede con due fasi:

1. Si prendono tutti gli $x \in S_1 : xy \in S_i$ e si mettono gli y in S_{i+1}
2. Si prendono tutti gli $x \in S_i : xy \in S_1$ e si mettono gli y in S_{i+1}

Al primo passaggio, $i = 1$ dunque i due insiemi coincideranno.

L'algoritmo termina in tre casi:

1. Uno degli S_i contiene una parola del codice. In questo caso il codice non è UD.
2. Si arriva ad avere un S_i vuoto. In questo caso il codice è UD.
3. Uno degli S_i è esattamente un set già visto in un'iterazione precedente S_j con $j < i$. In questo caso il codice è UD.

> Esempio
> $S_1 = \{A, E, C, ABB, CED, BBEC\}$
>
> Costruiamo S_2
> Fase A:
> A è prefisso di ABB, dunque inseriamo BB in S_2
> C è prefisso di CED, dunque inseriamo ED in S_2
>
> Fase B:
> A è prefisso di ABB, ma abbiamo già inserito BB in S_2
> C è prefisso di CED, ma abbiamo già inserito ED in S_2
>
> Dunque $S_2 = \{BB, ED\}$
>
> Ora confrontiamo S_1 e S_2 e costruiamo S_3
> Fase A:
> E è prefisso di ED, dunque inseriamo D in S_3
>
> Fase B:
> BB è prefisso di $BBEC$, dunque inseriamo EC in S_3
>
> Dunque $S_3 = \{D, EC\}$
>
> Ora confrontiamo S_1 e S_3 e costruiamo S_4
> Fase A:

> E è prefisso di EC, dunque inseriamo C in S_4
>
> Fase B:
> $\nexists$
>
> Dunque $S_4 = \{C\}$
> L'algoritmo termina perché S_4 contiene C, che è una parola del codice. Dunque il codice non è UD.

Un Codice Python per questo algoritmo è il seguente

```python
def SardinasPatterson(C: set, show=False):
    S = C
    Ss = []
    while (len(S) != 0):
        nextS = set()
        # Caso A
        for i in C:
            for j in S:
                if len(i) < len(j) and j[:len(i)] == i:
                    nextS.add(j[len(i):])
        # Caso B
        for i in S:
            for j in C:
                if len(i) < len(j) and j[:len(i)] == i:
                    nextS.add(j[len(i):])
        # Stampo a video il nuovo set
        if show:
            print(nextS)
        # Controllo se una parola del nuovo set è in C
        for i in nextS:
            if i in C:
                return False
        # Controllo se il nuovo set è già stato visto
        if nextS in Ss:
            return True
        else:
            Ss.append(nextS)
        S = nextS.copy()
    return True
```

Esercizio

Determinare se il codice $\{A, BCA, DE, CBC, AABC, C\}$ è univocamente decodificabile. Se non lo è, trovare una stringa non decodificabile.

$$S_1 = \{A, BCA, DE, CBC, AABC, C\}$$
$$S_2 = \{ABC, BC\}$$
$$S_3 = \{BC, A\}$$

A è una parola del codice, dunque il codice non è univocamente decodificabile.
Ad esempio, la stringa $CBCA$ potrebbe essere interpretata sia come C, BCA che come CBC, A