Teoria dell'Informazione

Simone Alessandro Casciaro

25 Ottobre 2024

Lezione 8: Derivati dell'Entropia

Nota: in questa lezione (e molto probabilmente anche fino alla fine del corso), considereremo $\boxed{d = 2}$. Dunque, si eviterà di indicare tale dato nelle definizioni delle entropie e dei logaritmi. In ogni caso, tutte le dimostrazioni presenti in questa lezione sono valide $\forall d > 1$.

Significato pratico dell'Entropia

Nella Lezione 4 abbiamo parlato dell'Entropia come indice di eterogeneità, ma cosa rappresenta l'entropia?
Nella `Teoria dell'Informazione`, L'Entropia $H(X)$ rappresenta la quantità di informazione da mandare per comunicare l'evento X e, dunque, il numero medio di simboli da spedire sul canale.
Parlando dell'Entropia Congiunta, $H(X, Y)$ rappresenta la quantità di informazione da mandare sul canale per comunicare sia X sia Y.
Infine, l'Entropia Condizionata $H(Y|X)$ rappresenta la quantità di informazione da mandare per comunicare Y sapendo che ho già mandato X in precedenza.

Intuitivamente, se posso esprimere Y come funzione di X, allora non ha senso mandare Y in quanto possiamo ricavarlo da X. Infatti, si ha questa relazione:

$$H(Y|X) = 0 \iff Y = g(X)$$

> Esempio
> $\mathbb{X} = \{-1, 0, 1\}$
>
> Indichiamo con X una variabile aleatoria e $Y = X^2$
> $H(Y|X) = 0$ perché, una volta spedito X sul canale, possiamo conoscere anche il risultato della variabile Y.

> Diversamente, $H(X|Y) \neq 0$ perché aver spedito Y sul canale non sempre è sufficiente a sapere il valore di X. Ad esempio, se $Y=1$, non sappiamo se $X=1$ oppure $X=-1$ e dunque abbiamo bisogno di un'informazione aggiuntiva.

Definizione di Entropia Condizionata

Ricordando che $p(x)=\sum_{y \in Y} p(x,y)$ e che $p(y|x)=\frac{p(x,y)}{p(x)}$, indichiamo l'entropia condizionata come segue:

$$\begin{aligned}
H(Y|X) &= \sum_{x \in \mathbb{X}} p(x) H(Y|X=x) \\
&= \sum_{x \in \mathbb{X}} p(x) \sum_{y \in \mathbb{Y}} p(y|x) \log \frac{1}{p(y|x)} \\
&= \sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{Y}} p(x) p(y|x) \log \frac{1}{p(y|x)} \\
&= \sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{Y}} p(x,y) \log \frac{1}{p(y|x)}
\end{aligned}$$

Quindi,

$$H(Y|X)=\sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{Y}} p(x,y) \log \frac{1}{p(y|x)}$$

Teorema: Chain Rule per l'Entropia

Tesi

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

Dimostrazione

$$
\begin{aligned}
H(X,Y) &= -\sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}} p(x,y)\log p(x,y) \\
&= -\sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}} p(x,y)\log\left(p(x)*p(y|x)\right) \\
&= -\sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}} p(x,y)\left[\log p(x)+\log p(y|x)\right] \\
&= -\sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}} p(x,y)\log p(x) - \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}} p(x,y)\log p(y|x) \\
&= -\sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}} p(x,y)\log p(x) + H(Y|X) \\
&= -\sum_{x\in\mathbb{X}} p(x)\log p(x) + H(Y|X) \\
&= H(X) + H(Y|X)
\end{aligned}
$$

Nota: Il risultato vale anche se condizioniamo il tutto a una terza variabile Z:

$$H(X,Y|Z) = H(X|Z) + H(Y|X,Z)$$

Informazione Mutua

L'informazione mutua è una misura che indica quanta informazione rilascia Y rispetto a X.

$$I(X,Y) = \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$$

L'informazione mutua è sempre ≥ 0.
Inoltre,

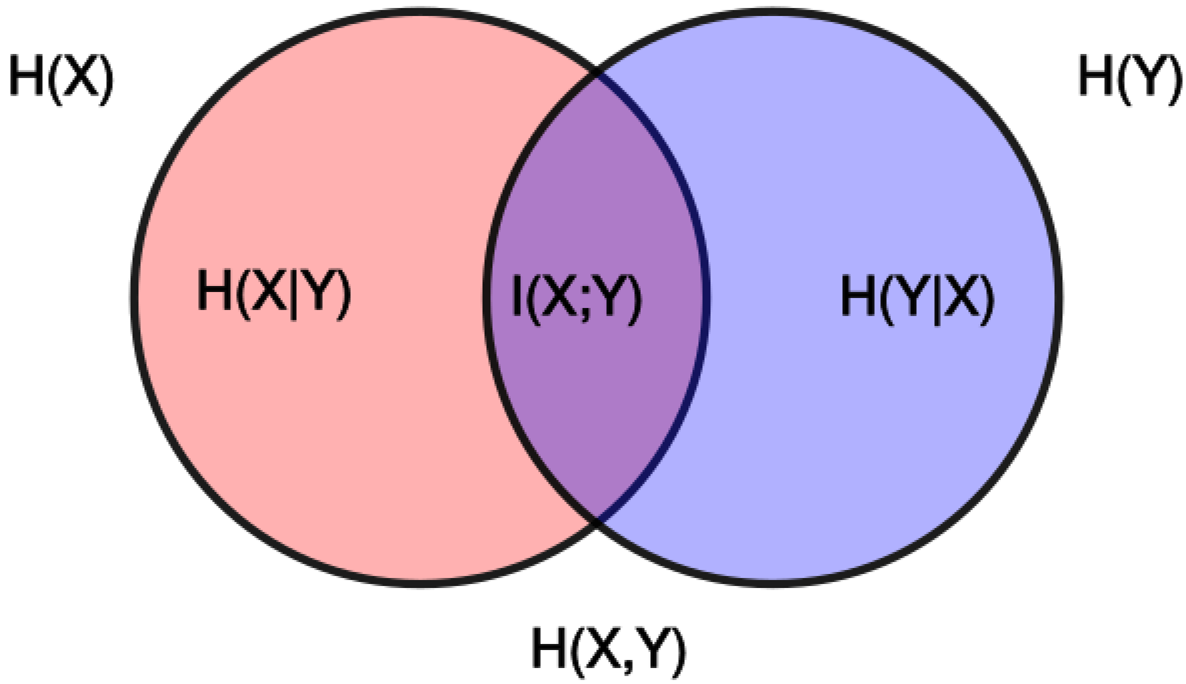
$$I(X,Y) = H(X) - H(X|Y)$$

Dimostrazione

$$
\begin{aligned}
I(X,Y) &= \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\
&= \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}} p(x,y)\log\frac{p(y)p(x|y)}{p(x)p(y)} \\
&= \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}} p(x,y)\log\frac{p(x|y)}{p(x)} \\
&= \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}} p(x,y)\left[\log p(x|y) - \log p(x)\right] \\
&= \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}} p(x,y)\log\frac{1}{p(x)} + \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}} p(x,y)\log p(x|y) \\
&= \sum_{x\in\mathbb{X}} p(x)\log\frac{1}{p(x)} - \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{Y}} p(x,y)\log\frac{1}{p(x|y)} \\
&= H(X) - H(X|Y)
\end{aligned}
$$

Ricapitolando le informazioni viste fino ad ora, si hanno le seguenti relazioni

$$
\begin{gathered}
I(X,Y) = H(X) + H(Y) - H(X,Y) \\
H(X) \geq H(X|Y) \\
H(Y) \geq H(Y|X) \\
H(X) = H(X|Y) + I(X,Y) \\
H(X,Y) = H(X) + H(Y) - I(X,Y)
\end{gathered}
$$



Informazione Mutua Condizionata

$$I(X, Y|Z) = \sum_{x \in \mathbb{X}} \sum_{y \in \mathbb{Y}} \sum_{z \in \mathbb{Z}} p(x, y, z) \log \frac{p(x, y, z)}{p(x|z)p(y|z)}$$

Teorema: Data Processing Inequality

Supponiamo di avere una sorgente $< \mathbb{X}, \mathbb{P} >$ e di spedire un messaggio sul canale, indicato con X. Se il canale è privo di rumore, allora il ricevente riceve esattamente X.

$$X \quad \overbrace{\longrightarrow}^{\text{Canale senza Rumore}} \quad X$$

Tuttavia, se invece il canale presenta del rumore, allora il messaggio ricevuto è Y. Supponiamo di voler migliorare Y tramite un algoritmo per ripulire il rumore, in modo tale da ricevere un nuovo messaggio Z.

$$X \quad \overbrace{\longrightarrow}^{\text{Canale con Rumore}} \quad Y \quad \overbrace{\longrightarrow}^{\text{Algoritmo per ripulire Rumore}} \quad Z$$

Ci chiediamo che relazione intercorre tra $I(X,Z)$ e $I(X,Y)$, ovvero se è possibile aumentare l'informazione rispetto a X usando Z al posto di Y.

Ipotesi

Siano X, Y, Z variabili aleatorie con codominio finito per la sorgente $< \mathbb{X}, \mathbb{P} >$ con $p(x,y,z)$ tale che $p(x,z|y) = p(x|y)p(z|y) \quad \forall x,y,z$ (Ovvero X, Z sono indipendenti dato Y)

Tesi

$$I(X,Y) \geq I(X,Z)$$

Dimostrazione

$$
\begin{aligned}
I(X,Y,Z) &= \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{X}}\sum_{z\in\mathbb{X}} p(x,y,z)\log\frac{p(x,y,z)}{p(x)p(y,z)} \\
&= \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{X}}\sum_{z\in\mathbb{X}} p(x,y,z)\log\frac{p(y|x,z)p(x,z)}{p(x)p(y,z)} \\
&= \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{X}}\sum_{z\in\mathbb{X}} p(x,y,z)\log\frac{p(y|x,z)p(x,z)}{p(x)p(y|z)p(z)} \\
&= \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{X}}\sum_{z\in\mathbb{X}} p(x,y,z)\left[\log\frac{p(x,z)}{p(x)p(z)} + \log\frac{p(y|x,z)}{p(y|z)}\right] \\
&= \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{X}}\sum_{z\in\mathbb{X}} p(x,y,z)\log\frac{p(x,z)}{p(x)p(z)} + \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{X}}\sum_{z\in\mathbb{X}} p(x,y,z)\log\frac{p(y|x,z)}{p(y|z)} \\
&= \sum_{x\in\mathbb{X}}\sum_{z\in\mathbb{X}} p(x,z)\log\frac{p(x,z)}{p(x)p(z)} + \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{X}}\sum_{z\in\mathbb{X}} p(x,y,z)\log\frac{p(y|x,z)}{p(y|z)} \\
&= I(X,Z) + \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{X}}\sum_{z\in\mathbb{X}} p(x,y,z)\log\left(\frac{p(y|x,z)}{p(y|z)}\frac{p(x|z)}{p(x|z)}\right) \\
&= I(X,Z) + \sum_{x\in\mathbb{X}}\sum_{y\in\mathbb{X}}\sum_{z\in\mathbb{X}} p(x,y,z)\log\frac{p(x,y|z)}{p(y|z)p(x|z)} \\
&= I(X,Z) + I(X,Y|Z)
\end{aligned}
$$

Dunque,

$$
I(X,Y,Z) = I(X,Z) = I(X,Y|Z) \tag{1}
$$

Tuttavia, se al posto di condizionare su Z lo avessimo fatto su Y, avremmo ottenuto

$$
I(X,Y,Z) = I(X,Y) + I(X,Z|Y) \tag{2}
$$

Possiamo eguagliare le equazioni 1 e 2, ottenendo

$$
I(X,Z) + I(X,Y|Z) = I(X,Y) + I(X,Z|Y)
$$

Dato che X e Z sono indipendenti se condizionate a Y, $I(X,Z|Y) = 0$ e dato che l'informazione mutua è sempre positiva, abbiamo dimostrato che

$$I(X,Y) \geq I(X,Z)$$

e quindi non è possibile aumentare l'informazione rispetto a quello effettivamente ricevuto dal canale.