Introducción a la Teoría de la Información

Información Mutua - Desigualdad de Fano

Facultad de Ingeniería, UdelaR

Agenda

Información Mutua

Definición

La *Información Mutua* entre dos variables aleatorias X, Y con distribución conjunta $p(x, y)$ se define como la divergencia entre la distribución conjunta y la distribución dada por el producto de las marginales $p(x)p(y)$.

$$\begin{aligned} I(X;Y) &= D(p(x,y)||p(x)p(y)) \\ &= \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \\ &= E_{p(x,y)}\log\frac{p(X,Y)}{p(X)p(Y)} \end{aligned}$$

La Información Mutua se expresa en bits.

La Información Mutua es No Negativa

Lema

$I(X;Y) \geq 0$ con igualdad si y sólo si X, Y son independientes.

Demostración.

De la definición, $I(X;Y) = D(p(x,y)||p(x)p(y)) \geq 0$ con igualdad si y sólo si $p(x,y) = p(x)p(y)$. $\square$

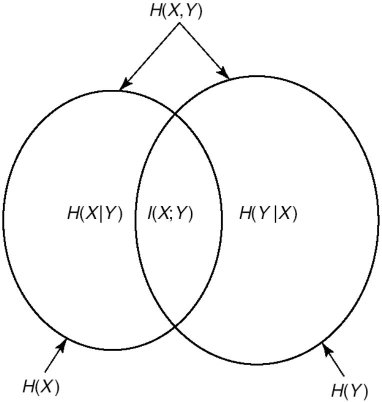
Relaciones entre Información Mutua y Entropía

$$\begin{aligned}
I(X;Y) &= D(p(x,y)||p(x)p(y)) \\
&= E_{p(x,y)} \log \frac{1}{p(X)p(Y)} - E_{p(x,y)} \log \frac{1}{p(X,Y)} \\
&= E_{p(x,y)} \left(\log \frac{1}{p(X)} + \log \frac{1}{p(Y)} \right) - E_{p(x,y)} \log \frac{1}{p(X,Y)} \\
I(X;Y) &= H(X) + H(Y) - H(X,Y)
\end{aligned}$$

$$\begin{aligned}
I(X;Y) &= H(X) + H(Y) - H(X,Y) \\
&= H(X) + H(Y) - (H(Y) + H(X|Y)) \\
I(X;Y) &= H(X) - H(X|Y) = H(Y) - H(Y|X) = I(Y;X)
\end{aligned}$$

$I(X;X) = H(X) - H(X|X) = H(X)$

Diagrama de Venn



$$
\begin{aligned}
I(X;Y) &= H(X) + H(Y) - H(X,Y) \\
&= H(X) - H(X|Y) \\
&= H(Y) - H(Y|X) \\
&= I(Y;X)
\end{aligned}
$$

Información Mutua Condicional

Definición

$$
\begin{aligned}
I(X;Y|Z) &= D(p(x,y|z)||p(x|z)p(y|z)) \\
&= E_{p(x,y,z)} \log \frac{p(X,Y|Z)}{p(X|Z)p(Y|Z)} \\
&= H(X|Z) - H(X|Y,Z)
\end{aligned}
$$

La Información Mutua Condicional es No Negativa

Lema

$I(X;Y|Z) \geq 0$ con igualdad si y sólo si X, Y son condicionalmente independientes dado Z.

Demostración.

De la definición, $I(X;Y|Z) = D(p(x,y|z)||p(x|z)p(y|z)) \geq 0$ con igualdad si y sólo si $p(x,y|z) = p(x|z)p(y|z)$ para todo z tal que $p(z) > 0$. □

Regla de la Cadena

Teorema

$I(X_1 \ldots X_n; Y) = \sum_{i=1}^{n} I(X_i; Y | X_{i-1}, \ldots X_1)$

Demostración.

$$
\begin{aligned}
I(X_1 \ldots X_n; Y) &= H(X_1 \ldots X_n) - H(X_1 \ldots X_n | Y) \\
&= \sum_{i=1}^{n} H(X_i | X_{i-1} \ldots X_1) - \sum_{i=1}^{n} H(X_i | X_{i-1} \ldots X_1, Y) \\
&= \sum_{i=1}^{n} \Big(H(X_i | X_{i-1} \ldots X_1) - H(X_i | X_{i-1} \ldots X_1, Y) \Big) \\
&= \sum_{i=1}^{n} I(X_i; Y | X_{i-1}, \ldots X_1)
\end{aligned}
$$

$\square$

Desigualdad de Procesamiento de Datos

Teorema

Si $X \to Y \to Z$, entonces $I(X;Y) \geq I(X;Z)$.

Demostración.

$$\begin{aligned} I(X;Y,Z) &= I(X;Z) + I(X;Y|Z) \\ &= I(X;Y) + \underbrace{I(X;Z|Y)}_{=0} \end{aligned}$$

$\square$

Corolario

Si $X \to Y \to Z$, entonces $I(X;Y|Z) \leq I(X;Y)$.

Desigualdad de Procesamiento de Datos

Corolario

En particular $I(X;Y) \geq I(X;f(Y))$

Demostración.

$X, Y, Z = f(Y)$ forman una cadena de Markov. $\square$

Concavidad/Convexidad de $I(X;Y)$

Teorema

Sea $(X,Y) \sim p(x,y) = p(x)p(y|x)$.

1. Dada $p(y|x)$ fija, $I(X;Y)$ es una función cóncava del vector de probabilidad $p(x)$.
2. Fijado el vector de probabilidad $p(x)$, $I(X;Y)$ es una función convexa de $p(y|x)$.

Desigualdad de Fano

Teorema

Sean X, Y variables aleatorias y $\hat{X} = f(Y)$ una estimación de X. Sea $P_e = P\{\hat{X} \neq X\}$

$$H(P_e) + P_e \log(|\mathcal{X}| - 1) \geq H(X|Y)$$

Corolario

$$1 + P_e \log(|\mathcal{X}| - 1) \geq H(X|Y)$$

$$P_e \geq \frac{H(X|Y) - 1}{\log(|\mathcal{X}| - 1)}, \quad |\mathcal{X}| > 2\,.$$

Desigualdad de Fano

Demostración.

Definimos $E = \begin{cases} 1 & \text{si } \hat{X} \neq X \\ 0 & \text{si } \hat{X} = X \end{cases}$

$$\begin{aligned} H(E,X|Y) &= H(X|Y) + \overbrace{H(E|X,Y)}^{=0} \\ &= \underbrace{H(E|Y)}_{\leq H(P_e)} + \underbrace{H(X|E,Y)}_{\leq P_e \log(|\mathcal{X}|-1)} \end{aligned}$$

La cota sobre $H(X|E,Y)$ surge de que, como $H(X|Y=y,E=0)=0$ para todo $y \in \mathcal{Y}$, entonces

$$\begin{aligned} H(X|E,Y) &= \sum_{y\in\mathcal{Y}} P(Y=y,E=1)H(X|Y=y,E=1) \\ &\leq \sum_{y\in\mathcal{Y}} P(Y=y,E=1)\log(|\mathcal{X}|-1) \\ &= P_e \log(|\mathcal{X}|-1) \end{aligned}$$

□