

Conceptos de entropía e información mútua

Abdelmalik Moujahid

Grupo de Inteligencia Computacional
Universidad del País Vasco UPV/EHU

2014-2015

Índice

- 1 Cantidad de información
- 2 Entropía de una variable
- 3 Información mútua
- 4 Ejercicios

Índice

1 Cantidad de información

2 Entropía de una variable

3 Información mútua

4 Ejercicios

Introducción

Dada una variable aleatorio X que toma valores en un conjunto denominado \mathcal{X} , y sea $P_X(x)$ su ley de probabilidad:

$$X \sim P_X(x)$$

$$P_r(X = x) = P_X(x), \forall x \in \mathcal{X}$$

Esta ley de probabilidad satisface las siguientes condiciones:

- $P_X(x) \geq 0, \forall x \in \mathcal{X}$
- $\sum_{x \in \mathcal{X}} P_X(x) = 1$

Definición

La cantidad de información de una variable aleatoria X es una medida de reducción de la incertidumbre asociada a los valores tomados por esta variable, y se define como:

$$I(X) = -\log_2 P_X(x), \forall x \in \mathcal{X}$$

- Si $P_X(x) \cong 1 \Rightarrow I(X) \cong 0$
- Si $P_X(x) \cong 0 \Rightarrow I(X) \cong +\infty$

Cuanto mas probable es un suceso menor cantidad de información aporta

Variable aleatoria binaria

Urna con 9 bolas negras y 1 bola blanca. Se efectúan extracciones sin reemplazamiento

- Se saca una bola blanca. Este suceso proporciona una alta información, ya que la incertidumbre sobre la siguiente extracción desaparece
- Se saca una bola negra. La información que proporciona este suceso es pequeña, ya que la incertidumbre acerca de la siguiente extracción se mantiene

Variable aleatoria definida en un conjunto finito

Al lanzar un dado si nos dicen que ha salido:

- un número menor que 2, más información (reduce más la incertidumbre) que
- un número múltiplo de 2

Índice

1 Cantidad de información

2 Entropía de una variable

3 Información mútua

4 Ejercicios

Definición

La entropía de una variable aleatoria X es una medida cuantitativa de la incertidumbre asociada a los valores tomados por esta variable, y se define como:

$$H(X) = - \sum_{x \in \chi} P_X(x) \log_2 P_X(x)$$

Se verifica

- $0 \leq H(X) \leq \log_2 n$, donde n es la cardinalidad del espacio χ
- $H(X) = 0 \Leftrightarrow \exists x$ con $P_X(x) = 1$
- Si X es una v.a uniforme discreta ($P_X(x) = \frac{1}{n} \forall x \in \chi$), entonces $H(X) = \log_2 n$

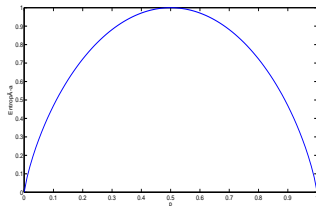
Nota: Si $P_X(x) = 0$, entonces $P_X(x) \log P_X(x) = 0$

Variable aleatoria binaria

Dada una variable aleatoria binaria $X \in \{0, 1\}$ con la ley de probabilidad $P_X(0) = p$ y $P_X(1) = 1 - p$, la entropía es:

$$H(X) = -p \log_2 p - (1 - p) \log_2 (1 - p)$$

- Es evidente que $H(p) = H(1 - p)$, y entonces $H(p)$ es una función simétrica centrada en $p = 0,5$.



Entropía conjunta

Sea $X \in \mathcal{X}$ y $Y \in \mathcal{Y}$ dos variables aleatorias con ley de probabilidad conjunta $P_{XY}(x, y)$, donde $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

La entropía conjunta viene dada por:

$$H(X, Y) = - \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} p(x, y) \log_2 p(x, y)$$

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log_2 p(x, y)$$

Entropía conjunta

Se verifica que,

$$H(X, Y) \geq H(X)$$

Esta desigualdad viene del hecho de que la probabilidad conjunta de dos eventos es siempre inferior o igual a la probabilidad de cada evento: $p(x, y) \leq p(x), \forall x \in \mathcal{X}, y \in \mathcal{Y}$:

$$\begin{aligned} H(X, Y) &= - \sum_{x, y} p(x, y) \log_2 p(x, y) \\ &= - \sum_{x, y} p(y|x)p(x) \log_2 p(y|x)p(x) \\ &\geq - \sum_{x, y} p(y|x)p(x) \log_2 p(x) = - \sum_x \left(\sum_y p(y|x)p(x) \right) \log_2 p(x) = H(X) \end{aligned}$$

Entropía condicional

Sea $X \in \mathcal{X}$ y $Y \in \mathcal{Y}$ dos variables aleatorias con ley de probabilidad conjunta $P_{XY}(x, y)$, donde $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

La entropía condicional de X dada Y es:

$$H(X|Y) = \sum_{y \in \mathcal{Y}} p(y) H(X|Y = y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log_2 p(x|y)$$

Demostrar la ley de entropías totales dada por:

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

Índice

1 Cantidad de información

2 Entropía de una variable

3 Información mútua

4 Ejercicios

Definición

La cantidad de información mútua entre dos v.a. X , Y mide la reducción en la incertidumbre en X cuando se conoce el valor de Y , y se define como:

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

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

Definición

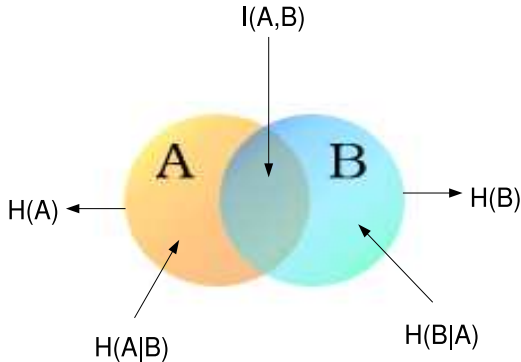


Figura: Diagrama de Venn. Caso de dos variables aleatorias

Propiedades

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

$$= - \sum_{i=1}^n p(x_i) \log_2 p(x_i) + \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 p(x_i|y_j)$$

$$= - \sum_{i=1}^n p(x_i) \log_2 p(x_i) + \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 p(x_i, y_j) \\ - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 p(y_j)$$

$$= \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 p(x_i, y_j) \\ - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) [\log_2 p(x_i) + \log_2 p(y_j)]$$

$$= \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) [\log_2 p(x_i, y_j) - (\log_2 p(x_i) + \log_2 p(y_j))]$$

$$= \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 \frac{p(x_i, y_j)}{p(x_i) \cdot p(y_j)}$$

Propiedades

Se verifica:

- $I(X, Y) = I(Y, X)$
- $I(X, Y|Z) = \sum_{k=1}^r p(z_k) I(X, Y|Z = z_k)$
$$= \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^r p(x_i, y_j, z_k) \log \frac{p(x_i, y_j | z_k)}{p(x_i | z_k) \cdot p(y_j | z_k)}$$
- $I(X, Y|Z) = H(X|Z) + H(Y|Z) - H(X, Y|Z)$
- $I(X, Y|Z) = 0 \iff X$ e Y son condicionalmente independientes dado Z
 - X e Y son condicionalmente independientes dado Z
 $\iff p(x|y, z) = p(x|z)$ para todo x, y, z

Índice

1 Cantidad de información

2 Entropía de una variable

3 Información mútua

4 Ejercicios

Ejercicio 1

- Tenemos un tetraedro y dos cubos. El tetraedro tiene sus caras numeradas del 1 al 4. El primer cubo del 1 al 6 y el segundo cubo tiene tres caras numeradas como 1 y las tres restantes como 2.
- Se escoge al azar uno de los tres objetos y se considera asimismo la cara expuesta de dicho objeto.
- Sean: X la v.a. denotando el objeto e Y la v.a. denotando la cara. Calcular: $H(X)$, $H(Y)$ y $H(Y|X)$

Ejercicio 2

- Tenemos dos monedas que denotamos por m_1 y m_2 . La moneda m_1 es equilibrada, mientras que en la moneda m_2 se tiene una probabilidad de 0,90 de obtener cara. Además contamos con un dado de caras numeradas del 1 al 6.
- Si el resultado al lanzar el dado es mayor que 4 se elige la moneda m_1 , en caso contrario se elige la moneda m_2 .
- Denotamos por X la variable aleatoria que representa la moneda escogida, por Y la variable aleatoria que recoge el resultado al lanzar la moneda (cara o cruz). Calcular: $H(X)$, $H(Y)$ e $I(X, Y)$.

Ejercicio 3

- Tenemos dos monedas: A en la cual la probabilidad de cara es $\frac{1}{2}$, y B con probabilidad de cara igual a 1.
- Se elige una moneda al azar, se lanza dos veces y se anota el número de caras.
- X v.a. denota la moneda escogida. Y v.a. denota el número de caras obtenidas. $H(X)$, $H(Y)$ y $I(X, Y)$

Ejercicio 4

- Se efectúan lanzamientos de una moneda hasta que aparezca una cruz. Sea X la variable aleatoria que indica el número de veces que se ha lanzado la moneda, y sea Y la variable aleatoria que indica la cara concreta de la moneda obtenida (cara o cruz). p es la probabilidad de obtener cruz en cada lanzamiento. Se sabe que el número de lanzamientos hasta la aparición de la primera cruz sigue una ley geométrica de parámetro n dada por:

$$p(X = n) = p(1 - p)^{n-1}; n = 1, 2, \dots, \infty$$

Ejercicio 4

- Demostrar que la entropía $H(X)$ de la variable aleatoria X es igual a $\frac{H(Y)}{p}$. Las siguientes expresiones pueden ser útiles:

$$\sum_{n=1}^{\infty} nr^n = \frac{r}{(1-r)^2}$$

$$\sum_{n=1}^{\infty} r^n = \frac{r}{(1-r)}$$

Conceptos de entropía e información mútua

Abdelmalik Moujahid

Grupo de Inteligencia Computacional
Universidad del País Vasco UPV/EHU

2014-2015