Teoría de la Información Estadística

Pedro Larrañaga, Iñaki Inza

Departamento de Ciencias de la Computación e Inteligencia Artificial

Universidad del País Vasco

`http://www.sc.ehu.es/isg/`

Índice

- Cantidad de información
- Entropía de una variable
- Divergencia de Kullback–Leibler
- Cantidad de información mútua

Cantidad de información

Cantidad de información asociada a un suceso

- 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

Cantidad de información

Cantidad de información como medida de reducción de la incertidumbre

- 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

Cantidad de información

- X variable aleatoria con posibles valores $x_1, \ldots x_n$ y probabilidades asociadas $p(x_1), \ldots, p(x_n)$
- $I(x_i) = -\log_2 p(x_i)$
  - Si $p(x_i) \approxeq 1 \Rightarrow I(x_i) \approxeq 0$
  - Si $p(x_i) \approxeq 0 \Rightarrow I(x_i) \approxeq +\infty$
- Cuanto mas probable es un suceso menor cantidad de información aporta

Entropía de una variable

- X variable aleatoria discreta con posibles valores $x_1, \ldots x_n$ y probabilidades asociadas $p(x_1), \ldots, p(x_n)$
- $I(X)$ variable aleatoria cantidad de información asociada a X, con posibles valores $I(x_1), \ldots I(x_n)$ y probabilidades asociadas $p(x_1), \ldots, p(x_n)$
- Se define la entropía de Shannon (1948), $H(X)$, de una variable aletoria discreta X como la esperanza matemática de la variable aleatoria asociada $I(X)$

$$H(X) = E(I(X)) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i)$$

- Si $p(x_i) = 0$, la indeterminación $p(x_i) \log_2 p(x_i)$ se resuelve asignándole el valor 0

Entropía de una variable

- X variable aleatoria de Bernouilli de parámetro p

$$P(X = x) = p^x (1-p)^{1-x} \quad x = 0, 1$$

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

  - Si $p = 0{,}50$ moneda correcta al aire

$$H(X) = -0{,}5 \log_2 0{,}5 - 0{,}5 \log_2 0{,}5 = 1$$

  - Si $p = 0{,}60$ moneda trucada al aire

$$H(X) = -0{,}6 \log_2 0{,}6 - 0{,}4 \log_2 0{,}4 = 0{,}97$$

  - Si $p = 0{,}90$ urna con 9 bolas negras y 1 blanca

$$H(X) = -0{,}9 \log_2 0{,}9 - 0{,}1 \log_2 0{,}1 = 0{,}468$$

Entropía de una variable

Se verifica

- $0 \leq H(X) \leq \log_2 n$
- $H(X) = 0 \iff \exists x_i$ con $p(x_i) = 1$
- Si X es variable aleatoria uniforme discreta, es decir $P(X = x_i) = \frac{1}{n}$ para todo $i = 1, \ldots, n$, entonces $H(X) = \log_2 n$

Entropía de una variable

- X v.a. con $x_1, \ldots, x_n$ y $p(x_1), \ldots, p(x_n)$
  Y v.a. con $y_1, \ldots, y_m$ y $p(y_1), \ldots, p(y_m)$
- (X, Y) v.a. bidimensional con $(x_1, y_1), \ldots, (x_1, y_m), \ldots, (x_n, y_1), \ldots, (x_n, y_m)$ y $p(x_1, y_1), \ldots, p(x_1, y_m), \ldots, p(x_n, y_1), \ldots, p(x_n, y_m)$
- $X|Y = y_j$ v.a. condicionada con $p(x_1|y_j), \ldots, p(x_n|y_j)$
- Entropía de la v.a. bidimensional conjunta (X, Y)

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

- Entropía de la v.a. X condicionada al valor $Y = y_j$

$$H(X|Y = y_j) = -\sum_{i=1}^{n} p(x_i|y_j)\log_2 p(x_i|y_j)$$

- Entropía de la v.a. X condicionada a la v.a. Y

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

Entropía de una variable

Ley de entropías totales: $H(X,Y) = H(X) + H(Y|X)$

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

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

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

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

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

$$= H(X,Y)$$

Entropía de una variable

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:

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

Entropía de una variable

Si X e Y son variables aleatorias independientes:

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

Divergencia de Kullback–Leibler

- Medir la distancia entre dos distribuciones de probabilidad –una de las cuales actua como referencia– definidas sobre la misma variable aleatoria X

$$D_{K-L}(p,q) = \sum_{i=1}^{n} p(x_i) \log \frac{p(x_i)}{q(x_i)}$$

- Se verifica que:
  - $D_{K-L}(p,q) \geq 0$
  - $D_{K-L}(p,q) = 0 \iff p(x_i) = q(x_i); i = 1, \ldots, n$
- En caso de que ninguna de las dos distribuciones de probabilidad se pueda considerar como distribución de referencia, se define la J–divergencia:

$$D_J(p,q) = D_{K-L}(p,q) + D_{K-L}(q,p) = D_J(q,p)$$

Cantidad de información mútua

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

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

Cantidad de información mútua

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

- $H(X) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$
- $H(X|Y)$
  - $P(Y=2|X=A) = \frac{1}{4}$; $P(Y=1|X=A) = \frac{1}{2}$; $P(Y=0|X=A) = \frac{1}{4}$; $P(Y=2|X=B) = 1$; $P(Y=1|X=B) = 0$; $P(Y=0|X=B) = 0$
  - $P(X=A, Y=0) = \frac{1}{8}$; $P(X=B, Y=0) = 0$; $P(X=A, Y=1) = \frac{1}{4}$; $P(X=B, Y=1) = 0$; $P(X=A, Y=2) = \frac{1}{8}$; $P(X=B, Y=2) = \frac{1}{2}$
  - $P(Y=0) = \frac{1}{8}$; $P(Y=1) = \frac{1}{4}$; $P(Y=2) = \frac{5}{8}$
  - $P(X=A|Y=0) = 1$; $P(X=B|Y=0) = 0$; $P(X=A|Y=1) = 1$; $P(X=B|Y=1) = 0$; $P(X=A|Y=2) = \frac{1}{5}$; $P(X=B|Y=2) = \frac{4}{5}$
  - $H(X|Y=2) = 0{,}7215$; $H(X|Y=1) = 0$; $H(X|Y=0) = 0$
  - $H(X|Y) = P(Y=0)\cdot H(X|Y=0) + P(Y=1)\cdot H(X|Y=1) + P(Y=2)\cdot H(X|Y=2) = \frac{5}{8}\cdot 0{,}7215 = 0{,}4509$
- $I(X,Y) = H(X) - H(X|Y) = 1 - 0{,}4509 = 0{,}5491$

Cantidad de información mútua

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

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

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

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

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

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

Cantidad de información mútua

Se verifica:

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