

# Teoría de la Información

## Fuentes de Información



UNIVERSIDAD DE VALLADOLID

Técnicas de Aprendizaje Automático

Grado de Ingeniería Informática

Doble Grado con Estadística (INDAT)

Departamento de Informática (ATC, CCIA y LSI)

# Transmisión de la Información I

Tiempo Oviedo	Probabilidad	Código A
Soleado	1/4	00
Nublado	1/4	01
Lluvia	1/4	10
Niebla	1/4	11

Longitud media por mensaje:

$$L = 2 \text{ binits/mensaje}$$

bit: unidad de información  
Binit: dígito binario

# Transmisión de la Información II

Tiempo Almería	Probabilidad	Código A	Código B
Soleado	1/4	00	10
Nublado	1/8	01	110
Lluvia	1/8	10	111
Bruma	1/2	11	0

$$L = 2P(S) + 3P(N) + 3P(L) + 1P(B) = \frac{14}{8} \text{ binits/mensaje}$$

Ahorro: 0.25  
binits/mensaje  
Ganancia: 12.5%

# Ejercicio 1

- Basado en la codificación anterior, se recibe esta tira de bit:

0001010010010101110

- Descodificar y calcular la longitud media de estos mensajes expresados en binit.
- Compare esta cantidad con la medida en la dispositiva anterior e intente explicar la diferencia



---

# Solución

---

- 0001010010010101110
- 0-0-0-10-10-0-10-0-10-10-111-0
- Bruma, bruma, bruma, soleado, soleado, bruma, soleado, bruma, soleado, bruma,, bruma, lluvia, bruma
- $L=19/12=1,5833$  binits/mensaje
- $L(\text{media})=14/8= 1,75$  binits/mensaje
- Las probabilidades a priori no se verifican en esta muestra tan pequeña de mensajes.

# Cantidad de Información

- Cantidad de información asociada a un suceso E.

$$I(E) = \log_a \frac{1}{P(E)}$$

- Según esta dependencia funcional, a menor probabilidad del suceso, mayor cantidad de información y viceversa.
- Dado que los mensajes usan algún código binario, el logaritmo suele ser en base 2

- La unidad de  $I(E)$  es en bits.  $\left\{ \begin{array}{l} I(\text{niebla} - \text{Oviedo}) = \log_2 \frac{1}{2^{-2}} = 2 \text{ bits} \\ I(\text{bruma} - \text{Almería}) = \log_2 \frac{1}{2^{-1}} = 1 \text{ bit} \end{array} \right\}$

# Fuente de Información sin memoria

$$S = \{s_1, s_2, s_3, \dots, s_q\} \longrightarrow s_i, s_j, \dots \qquad I(s_i) = \log \frac{1}{P(s_i)}$$

$$\text{Memoria nula} \Leftrightarrow P(s_i / s_j) = P(s_i) \quad \forall i, j = 1, 2, \dots, q \quad (i \neq j)$$

Cantidad de información media asociada a la fuente.

$$\sum_{i=1}^q P(s_i) I(s_i) = H(S)$$

Def: **Entropía** asociada a la fuente



# Entropía

$$S(\text{Oviedo}) = \{\text{soleado}, \text{nublado}, \text{lluvia}, \text{niebla}\} \Rightarrow H(S) = 2 \text{ bits}$$

$$S(\text{Almeria}) = \{\text{soleado}, \text{nublado}, \text{lluvia}, \text{bruma}\}$$

$$H(S) = \frac{1}{4} 2 + \frac{1}{8} 3 + \frac{1}{8} 3 + \frac{1}{2} 1 = \frac{14}{8} \text{ bits}$$

Coincide con la longitud media de los mensajes en este caso particular, pero esto no se puede generalizar (\*)

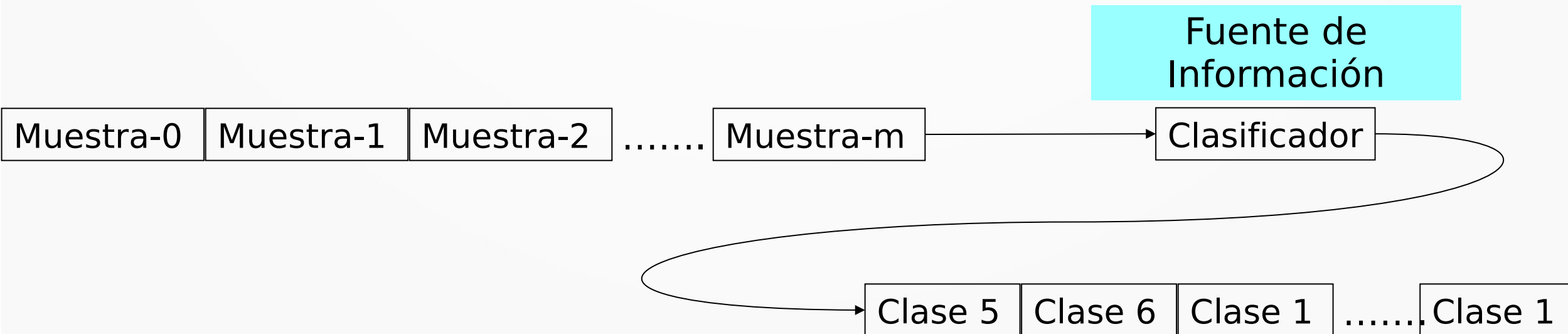
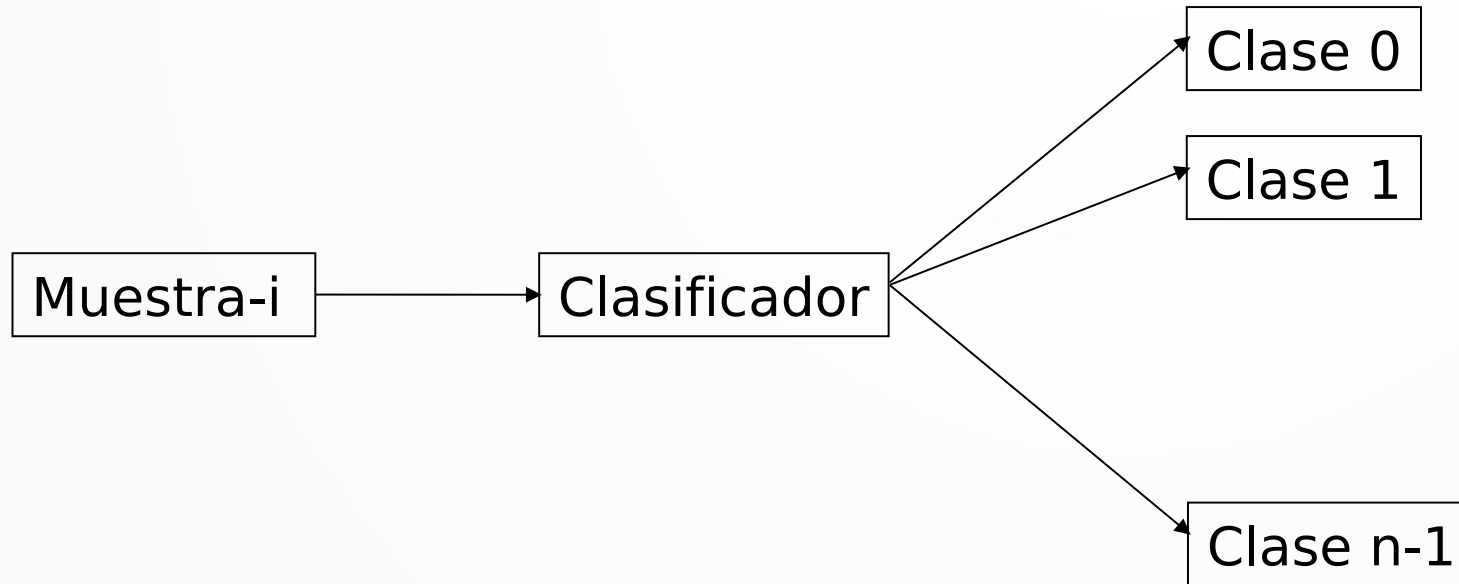
*Cuanto más entropía, más uniforme la distribución de los mensajes.*

$$S = \{c_1, c_2, \dots, c_n\} \Rightarrow H(S) = \sum_{i=1}^n P(c_i) I(c_i)$$

(\*)  $H(S)$  no depende de la codificación, pero sí de la probabilidad a priori



# Clasificación



# Clasificador como Fuente de Información

- En la salida del clasificador será:
  - Colección de etiquetas posibles :  $S = \{c_1, c_2, \dots, c_n\}$
  - Entropía asociada al clasificador:  $H(S) = \sum_{i=1}^n P(c_i) I(c_i)$
- Cada muestra  $x$  lleva asociada su etiqueta  $c(x)$
- Habrá experimentos donde todas las clases tienen el mismo número de muestras (equiprobables) pero, en general, no tiene por qué ser así.

# Ejemplo de clasificador y entropía

- Se tiene tabulada una serie de muestras, una por fila, cada una de las cuales viene dada por cuatro atributos (características) y una clase (Play Tennis)

Outlook	Temperature	Humidity	Wind	PlayTennis
Sunny	Hot	High	Weak	NO
Sunny	Hot	High	Strong	NO
Overcast	Hot	High	Weak	YES
Rain	Mild	High	Weak	YES
Rain	Cool	Normal	Weak	YES
Rain	Cool	Normal	Strong	NO
Overcast	Cool	Normal	Strong	YES
Sunny	Mild	High	Weak	NO
Sunny	Cool	Normal	Weak	YES
Rain	Mild	Normal	Weak	YES
Sunny	Mild	Normal	Strong	YES
Overcast	Mild	High	Strong	YES
Overcast	Hot	Normal	Weak	YES
Rain	Mild	High	Strong	NO

- Calcular la entropía del clasificador



---

# Solución

---

Clase	Probabilidad	$I(c)$
No	5/14	$\log_2(14/5)=1,4854$
Yes	9/14	$\log_2(14/9)=0,6734$
H(S)		0,940286 bits

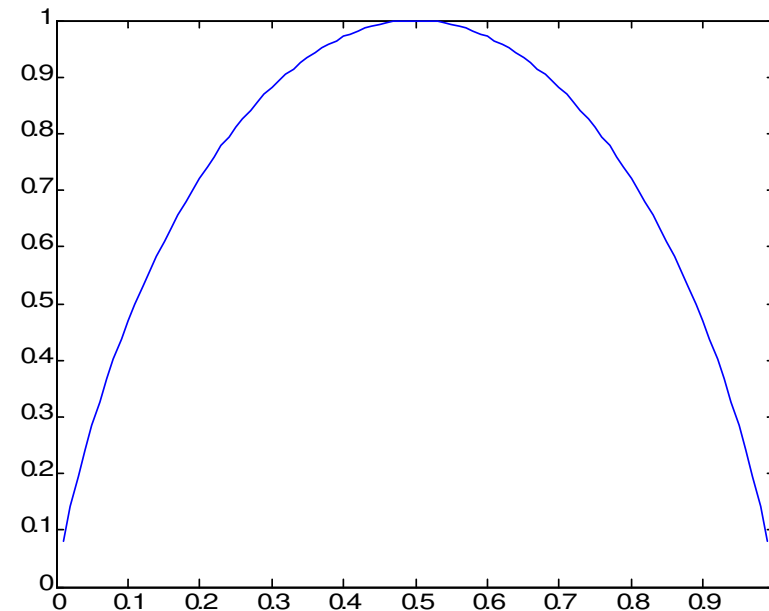


Ej: en python, la Entropía de un Clasificador Binario

# Clasificador binario

$$S = \{+, -\} \quad P(+) = x; \quad P(-) = 1 - x$$

$$H(S) = -x \log(x) - (1 - x) \log(1 - x)$$



# Índice Gini

- Sustituye a la entropía
- Resuelve los casos asintóticos en los que hay que aplicar el límite
- Es muy empleada en las bibliotecas de python, aunque ofrecen las dos posibilidades (Gini vs Entropy)

$$Gini(S) = 1 - \sum_{i=1}^n p(c_i)^2$$





# Ganancia de Información

---

- Reducción de la entropía por causa de la partición de los ejemplos debido a un atributo

$$G(S, A) = H(S) - \sum_{i=1}^k \frac{|S_{v_i}|}{|S|} H(S_{v_i})$$

$|S_{v_i}| = \text{instancias con } A = v_i$

- Ejercicio 3: Considérese una clasificación binaria de 14 ejemplos denotados por: 9+, 5-. Si se escoge un atributo  $A = \{bajo, alto\}$ , también binario, su distribución sería:

$bajo \leftarrow \{6+, 2-\}$

$alto \leftarrow \{3+, 3-\}$

Calcular  $G(S, A)$

# Entrega

- Hallar la ganancia en información usando la entropía para cada uno de los cuatro atributos de la última tabla (PlayTennis). O si se prefiere, del fichero de datos proporcionado por weka bajo el nombre de “weather nominal”.
- Se hará en un fichero de python (jupyter), en el que se calculará de manera automática:
  - Ganancia en Información respecto a cada atributo
  - Lo mismo, pero aplicando el índice GINI