

- Existen muchas otras alternativas basadas en regresiones, covarianzas o bien otros criterios de información para seleccionar atributos, a saber:
- **La «F-Regression»:** En el caso de la regresión F, se regresa cada variable individualmente y se determinan los valores F y p-value de cada regresión para posteriormente rankear y elegir las de mayor test F o bien menores p-values
- **Mutual Information:** Intuitivamente, la información mutua media mide la información que X e Y comparten, o de otro modo, mide en cuánto el conocimiento de una variable reduce nuestra incertidumbre sobre la otra. Por ejemplo, si X e Y son independientes, entonces conocer X no da información sobre Y y viceversa, por lo que su información mutua es cero. En el otro extremo, si X e Y son idénticas entonces toda información proporcionada por X es compartida por Y , es decir, saber X determina el valor de Y y viceversa. Por ello, la información mutua media es igual a la información contenida en Y (o X) por sí sola, también llamada la entropía de Y (o X : claramente si X e Y son idénticas tienen idéntica entropía).
- La información mutua media cuantifica la dependencia entre la distribución conjunta de X e Y y la que tendrían si X e Y fuesen independientes. La información mutua media es una medida de dependencia en el siguiente sentido: $I(X; Y) = 0$ si y solo si X e Y son variables aleatorias independientes. Esto es fácil de ver en una dirección: si X e Y son independientes, entonces $p(x,y) = p(x) p(y)$

- Ejemplo en Python:

```
import numpy as np
from sklearn.feature_selection import f_regression, mutual_info_regression

y=x.iloc[:,0]
x=x.iloc[:,1:]

f_test, _ = f_regression(x, y)
f_test /= np.max(f_test)

mi = mutual_info_regression(x, y)
mi /= np.max(mi)
```

f_test - Arreglo de NumPy	
	0
0	1.000
1	0.322
2	0.153
3	0.091
4	0.050
5	0.029
6	0.015
7	0.013

mi - Arreglo de NumPy	
	0
0	1.000
1	0.791
2	0.717
3	0.596
4	0.454
5	0.444
6	0.379
7	0.388

Entropía $H(X)$

We first introduce the concept of *entropy*, which is a measure of the uncertainty of a random variable. Can also be understood as the (im)purity of an arbitrary collection of examples (random variable).

Definition The *entropy* $H(X)$ of a discrete random variable X is defined by

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x). \quad (2.1)$$

Día	Cobre Sube	Dólar Baja
1	si	si
2	no	no
3	no	si
4	si	si
5	no	si
6	si	no
7	si	si
8	no	no
9	si	si
10	si	si
11	si	si
12	no	si
13	si	si
14	no	no

Distribución de Prob de Cobre Sube

Caso	Cuenta	Prob	log(Prob)
si	8	57%	-0.80735
no	6	43%	-1.22239
Total	14	100%	
H(Cobre Sube)	0.99		

Distribución de Prob de Dólar Baja

Caso	Cuenta	Prob	log(Prob)
si	10	71%	-0.48543
no	4	29%	-1.80735
Total	14	100%	
H(Dólar Baja)	0.86		

Entropía Conjunta $H(X,Y)$

Definition The *joint entropy* $H(X, Y)$ of a pair of discrete random variables (X, Y) with a joint distribution $p(x, y)$ is defined as

$$H(X, Y) = - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y), \quad (2.8)$$

Distr Prob Conjunta Cobre Sube y Dólar Baja

Caso	Cobre Sube	Dólar Baja	Cuenta	Prob	log(Prob)
Cobre Sube=no y Dólar Baja=no	no	no	3	21%	-2.22
Cobre Sube=si y Dólar Baja=no	si	no	1	7%	-3.81
Cobre Sube=no y Dólar Baja=si	no	si	3	21%	-2.22
Cobre Sube=si y Dólar Baja=si	si	si	7	50%	-1.00
			14		

Entropia Conjunta de Cobre Sube y Dólar Baja
 $H(\text{Cobre Sube}; \text{Dólar Baja})$

1.72

Entropía Condicional $H(Y/X)$

Definition [\[edit \]](#)

If $H(Y|X = x)$ is the entropy of the variable Y conditioned on the variable X taking a certain value x , then $H(Y|X)$ is the result of averaging $H(Y|X = x)$ over all possible values x that X may take.

Given discrete random variables X with Image \mathcal{X} and Y with Image \mathcal{Y} , the conditional entropy of Y given X is defined as: (Intuitively, the following can be thought as the weighted sum of $H(Y|X = x)$ for each possible value of x , using $p(x)$ as the weights)^[1]

$$\begin{aligned} H(Y|X) &\equiv \sum_{x \in \mathcal{X}} p(x) H(Y|X = x) \\ &= - \sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x) \\ &= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x) \\ &= - \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x, y) \log p(y|x) \\ &= - \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)}. \\ &= \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x, y) \log \frac{p(x)}{p(x, y)}. \end{aligned}$$

Entropía Condicional $H(Y/X)$

Distr Prob Conjunta Dólar Baja Cobre Sube

(Cobre sube)

Caso	Cobre Sube	Dólar Baja	Cuenta	Prob(x,y)	Prob(x)	Prob(x)/Prob(x,y)	log[Prob(x)/Prob(x,y)]
Cobre Sube=no y Dólar Baja=no	no	no	3	21%	42.86%	2.00	1.00
Cobre Sube=si y Dólar Baja=no	si	no	1	7%	57.14%	8.00	3.00
Cobre Sube=no y Dólar Baja=si	no	si	3	21%	42.86%	2.00	1.00
Cobre Sube=si y Dólar Baja=si	si	si	7	50%	57.14%	1.14	0.19
			14	Entropia condicional Dólar Baja dado Cobre Sube			0.74

Distr Prob Conjunta Cobre Sube Dólar Baja

(Dólar Baja)

Caso	Cobre Sube	Dólar Baja	Cuenta	Prob(x,y)	Prob(y)	Prob(y)/Prob(x,y)	log[Prob(x)/Prob(x,y)]
Cobre Sube=no y Dólar Baja=no	no	no	3	21%	28.57%	1.33	0.42
Cobre Sube=si y Dólar Baja=no	si	no	1	7%	28.57%	4.00	2.00
Cobre Sube=no y Dólar Baja=si	no	si	3	21%	71.43%	3.33	1.74
Cobre Sube=si y Dólar Baja=si	si	si	7	50%	71.43%	1.43	0.51
			14	Entropia condicional Cobre Sube dado Dólar Baja			0.86

Información Mutua $I(X,Y)$

Definition: The mutual information $I(X;Y)$ between the random variables X and Y is given by

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

Theorem 2.4.1 (Mutual information and entropy)

$$I(X;Y) = H(X) - H(X|Y) \quad (2.43)$$

$$I(X;Y) = H(Y) - H(Y|X) \quad (2.44)$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \quad (2.45)$$

$$I(X;Y) = I(Y;X) \quad (2.46)$$

$$I(X;X) = H(X). \quad (2.47)$$

$$\begin{aligned} &= \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)} \\ &= E_{p(x,y)} \left[\log_2 \frac{p(X,Y)}{p(X)p(Y)} \right] \end{aligned}$$

Distr Prob Conjunta Cobre Sube y Dólar Baja

(Cobre sube)

(Dólar Baja)

Caso	Cobre Sube	Dólar Baja	Cuenta	Prob(x,y)	Prob(x)	Prob(y)	Prob(x,y)/P(x)P(y)	log[Prob(x,y)/P(x)P(y)]
Cobre Sube=no y Dólar Baja=no	no	no	3	21.4%	42.9%	28.6%	1.75	0.81
Cobre Sube=si y Dólar Baja=no	si	no	1	7.1%	57.1%	28.6%	0.44	-1.19
Cobre Sube=no y Dólar Baja=si	no	si	3	21.4%	42.9%	71.4%	0.70	-0.51
Cobre Sube=si y Dólar Baja=si	si	si	7	50.0%	57.1%	71.4%	1.23	0.29
			14	Information Gain (o Mutal Information)			$I(\text{Dólar Baja}; \text{Cobre Sube})$	
								0.12

Entropía Dólar Baja: $H(\text{Dólar Baja})$

0.86

$H(\text{Dólar Baja} | \text{Cobre Sube})$

0.74

Information Gain (o Mutal Information)

$I(\text{Dólar Baja}; \text{Cobre Sube})$

0.12

Ratio de Ganancia de Información IGR

Problemas conocidos de la Métrica de Información Mutua:

- Puede sesgarse hacia variables que cuenten con muchos posibles valores únicos
 - caso típico atributo ID
- Para corregir este problema, nace el IGR “Information Gain Ratio”
- Se entiende como una normalización, o una penalización de la Información Mutua para aquellos atributos que cuenten con muchos posibles valores únicos.
- Se calcula dividiendo la Información Mútua entre X e Y, por la Entropía de X:

$$IGR = \frac{I(X,Y)}{H(X)}$$