Machine Learning

Introduction to Information Theory

Fernando Rodríguez Sánchez

Computational Intelligence Group

Universidad Politécnica de Madrid

27/01/2020

Table of contents

Table of contents

Entropy of a variable

X is a categorical variable with

- A set of possible values $x_1, \ldots, x_n$
- A set of associated probabilities $p(x_1), \ldots, p(x_n)$

Information of an event

Information provided by the event x_i: $I(x_i = -log_2\, p(x_i)$

- If $p(x_i) \cong 1 \Rightarrow I(x_i \cong 0)$
- If $p(x_i) \cong 0 \Rightarrow I(x_i \cong +\infty)$

A **very probable** event provides **little information**

Why the logarithm?

Entropy of a variable

Definition

The Shannon entropy $H(X)$ of the discrete random variable X is defined as its expected information value:

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

We assume $p(x_i)\ log_2\ p(x_i) = 0$ if $p(x_i) = 0$

Example

Assume X is a Bernouilli variable with parameter p

$$f(x) = p^x(1-p)^{1-x} \text{ where } x = \{0, 1\}$$

Its entropy would then be:

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

- If $p = 0.5$: $H(X) = (-0.5\ log_2\ 0.5) + (-0.5\ log_2\ 0.5) = 1$
- If $p = 0.6$: $H(X) = (-0.6\ log_2\ 0.6) + (-0.4\ log_2\ 0.4) = 0.97$
- If $p = 0.9$: $H(X) = (-0.9\ log_2\ 0.9) + (-0.1\ log_2\ 0.1) = 0.468$

Entropy of two variables

X is a categorical variable with

- $x_1, \ldots, x_n$
- $p(x_1), \ldots, p(x_n)$

Y is a categorical variable with

- $y_1, \ldots, y_m$
- $p(y_1), \ldots, p(y_m)$

(X,Y) bidimensional with

- $(x_1, y_1), \ldots, (x_1, y_m), \ldots, (x_n, y_1), \ldots, (x_n, y_m)$
- $p(x_1, y_1), \ldots, p(x_1, y_m), \ldots, p(x_n, y_1), \ldots, p(x_n, y_m)$

Entropy of two variables

Entropy of the bidimensional variable (X, Y)

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

Entropy of variable X conditioned to the value $Y = y_j$

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

Entropy of variable X conditioned to variable Y

$$H(X \mid Y) = \sum_{j=1}^{m} p(y_j)\ H(\mid Y = y_j)$$

Entropy of two variables

Total entropies law

$$H(X,Y) = H(X \mid Y) + H(Y)$$

If X and Y are **independent**:

- $H(X \mid Y) = H(X)$
- $H(Y \mid X) = H(Y)$
- $H(X,Y) = H(X) + H(Y)$

Table of contents

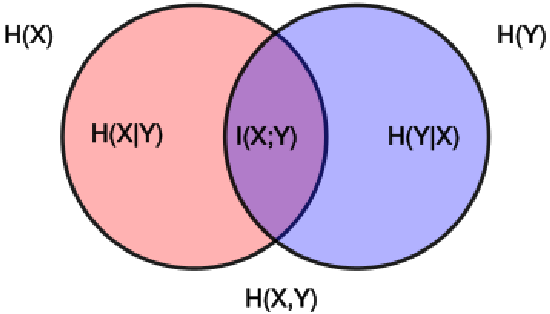
Mutual information

Definition

The mutual information between two variables X, Y measures the reduction in the uncertainty of variable X once the value of Y is known

$$I(X, Y) = H(X) - H(X \mid Y)$$

Mutual information



Machine Learning

Introduction to Information Theory

Fernando Rodríguez Sánchez

Computational Intelligence Group

Universidad Politécnica de Madrid

27/01/2020