6. Mutual Information

15 May 2023 14:50

Mutual Information (MI) is a measure of the dependency between two variables. It quantifies the amount of information obtained about one random variable through observing the other random variable. It is a fundamental quantity in information theory.

$$MI = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \left[\frac{p(x,y)}{p(x)\,p(y)} \right]$$

p(y)

where

p(x,y) → Joint prob of x and y

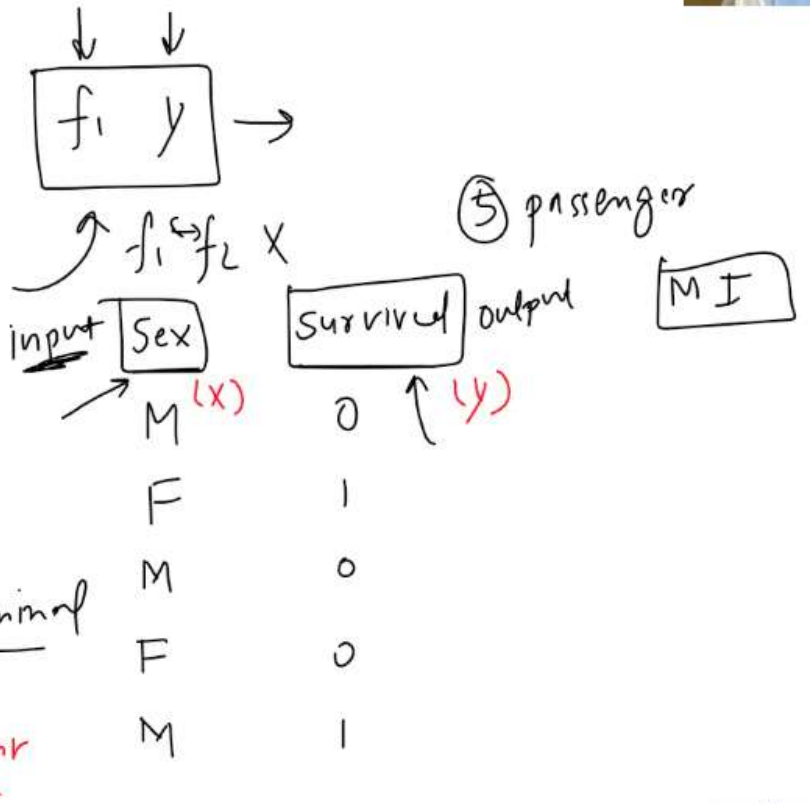
p(x) → marginal prob of x

p(y) → marginal prob of y

p(x,y) p(x) → marginal

P(x,y) → joint

P(x) P(y) → marginal



1. Joint Probability: This is the probability of two (or more) simultaneous events. For example, if we have two random variables, X and Y, the joint probability of X and Y is denoted as P(X, Y), and it represents the probability that X takes on a specific value and Y takes on a specific value at the same time. In other words, it represents the probability of both events happening at the same time.

2. Marginal Probability: This is the probability of an event occurring regardless of the

learnwith.campusx.in
j.s.aniket@gmail.com
+918180865500

OneNote for Windows 10

Home Insert Draw View Help

Shapes Ink to Shape Ink to Text

ependency between two variables. It
d about one random variable through
ndamental quantity in information theory.

$g\left[\frac{p(x,y)}{p(x)\,p(y)}\right]$

p(y)

p(x,y) p(x)

x and y

f x

f y

p(x,y) → joint

p(x) p(y) → marginal

f two (or more) simultaneous events. For
s, X and Y, the joint probability of X and Y is
probability that X takes on a specific value
e time. In other words, it represents the
he same time.

ty of an event occurring regardless of the

f1 y →

f1 ↪ f2 X

input Sex (x) | Survived (y) output

Sex	Survived
M	0
F	1
M	0
F	0
M	1

⑤ passenger

MI

m → 0 2/5

p(X=m AND Y=1)

p(X=m, Y=0)

sex survived

Sex \ Survived	0	1	
M	2/5	1/5	3/5
F	1/5	1/5	2/5
	3/5	2/5	

$[\ p(x)\ p(y)\ ]$ $\log$

$p(x,y)$ $p(x)$ → marginal

of X and Y

b of X $p(x,y)$ → joint

b of Y $p(x)\ p(y)$ → marginal

	(x)	(y)
M		0
F		1
M		0
F		0
M		1

Sex

	0	1	
M	2/5	1/5	3/5
F	1/5	1/5	2/5
	3/5	2/5	

ty of two (or more) simultaneous events. For ables, X and Y, the joint probability of X and Y is the probability that X takes on a specific value same time. In other words, it represents the at the same time.

ability of an event occurring regardless of the e two random variables, X and Y, the marginal P(X), and it represents the probability that X of the values of Y. The term "marginal" refers to g over the distribution of the other variable(s) ble of interest.

MI

$$\left[\frac{2}{5}\log\left(\frac{2/5}{3/5\times 3/5}\right)+\frac{1}{5}\log\left(\frac{1/5}{3/5\times 2/5}\right)+\frac{1}{5}\log\left(\frac{1/5}{3/5\times 2/5}\right)+\frac{1}{5}\log\left(\frac{1/5}{2/5\times 3/5}\right)\right]$$

Sex → Survived

s that make it useful for feature selection:

r positive, with zero indicating that the formation about one variable can be obtained

to obtain the distribution of the variable of interest.

Survived

Mutual Information has several properties that make it useful for feature selection:

1. It is non-negative: MI is always zero or positive, with zero indicating that the variables are independent (i.e., no information about one variable can be obtained by observing the other variable).
2. It is symmetric: MI(X, Y) = MI(Y, X). The mutual information from X to Y is the same as from Y to X.
3. It can capture any kind of statistical dependency: Unlike correlation, which only captures linear relationships, mutual information can capture any kind of relationship, including nonlinear ones.

How to deal with numerical variables