

Mutual Information \Rightarrow Filter Based

Mutual Information (MI) is a measure of the dependency between two variable. 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.

Formula:

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

where

$P(x, y) \rightarrow$ joint probability of x and y

$P(x) \rightarrow$ marginal prob of x

$P(y) \rightarrow$ marginal prob of y

example:-

input (x) Sex Survived (y) output

M	0
F	1
M	0
F	0
M	1

* Make contingency table

No. of blocks

		Survived (y)		
		0	1	Total
(x) Sex	M	① 2/5 $P(x,y)$	② 1/5 $P(x,y)$	3/5 $P(x)$
	F	③ 1/5 $P(x,y)$	④ 1/5 $P(x,y)$	2/5 $P(x)$
Total		3/5 $P(y)$	2/5 $P(y)$	

Marginal

$P(x=M, y=0)$
 Sex Survived

~~first~~ Marginal

use formula:-

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

The formula has to be used as many times as there are blocks. → block No. 2

block No. 1

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

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

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

→ block No. 3

→ block No. 4

$$MI = \text{Sex} \xrightarrow{\text{Importance}} \text{Survived}$$

- * $MI \uparrow \uparrow$ when $P(x, y) \uparrow \uparrow$
- * $MI \downarrow \downarrow$ when $P(x) P(y) \uparrow \uparrow$

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. No information about one variable can be obtained ~~from~~ by observing the other variable.
2. It is symmetric: $MI(X, Y) = MI(Y, X)$. The mutual information from X to Y is same as from Y to X .
3. It can capture any kind of statistical dependency: - Unlike correlation, which only captures linear relationship, Mutual Information captures any kind of relationship, including Non-linear.

Sex - Survived \rightarrow Linear \rightarrow Chi square
 \downarrow
 Capture

- * MI method apply on Numerical data but Chi square not.

example:-

age

survived

35	→ #2
60	→ #3
52	→ #3
16	→ #1
27	→ # (2)

change
to this
Number

0
1

1 change
1

Categorical

0

And then Apply MI
method

create histogram

↓

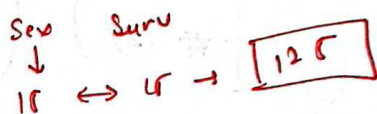
0-20
#1

21-50
#2

51-80
#3

Disadvantage

1. Estimation Difficulty: Estimating MI from data can be challenging, especially when the dimensionality of the data is high or the number of samples is low. This is because MI estimation often relies on techniques like binning or density estimation, which can be sensitive to the chosen parameters or assumptions.



2. Assumes Large Sample Sizes: MI works best with large sample sizes. With smaller sample sizes, the estimates of MI can be noisy and less reliable, which might lead to incorrect conclusions about the dependencies between variables.

3. Computationally Intensive: Calculating MI for many features can be computationally expensive, especially for continuous variables. This might be problematic for large datasets or for applications where computational resources or time are limited.

4. Difficulty with Continuous Variable: While MI theoretically applies to continuous variables, in practice it's often difficult to estimate MI between continuous variables due to the need for accurate density estimates. MI between continuous variables

due to the need for accurate density estimate, which is a ~~as~~ challenging problem in its own right.

↳ binning

5. No Direct Indication of the Nature of Relationships:

Although MI can identify the existence of a relationship between variables, it does not provide direct information about the nature of this relationship. (eg. linear, quadratic, etc.) This contrast with methods such as correlation, which directly indicates the strength and direction of a linear relationship.

MI \uparrow \rightarrow relation ✓
but not provide nature
linear, quadratic

6. Doesn't Account for Redundancy: Mutual information

measure the relevance of individual features to the target variables, but it doesn't take into account the redundancy among features. They might individually have high MI with the target, but if they are highly correlated, they might not provide much unique information. This can lead to the selection of redundant features.



what if strong relationship between f_1 f_2
multicollinearity.