Information Theory

Study notes

Shuqi XIAO

May 21, 2024

Outline

Basic Concepts

Self-information: Motivation

What is information?

How can we define "informational value" of "messages"?

- Messages are defined as **events** in Probability Theory.
- Informational value is defined by **three axioms**.

Why is it called 'self-information' instead of a simple 'information'? Stay tuned.

Self-information: Definition (I)

1. A certain event carries no information.
2. The less probable an event is, the more information it yields.
3. If two independent events occur, the total information is the sum of each.

Definition (the self-information of an event)

For an event X with the probability measure P, the self-information is

$$I(X) := -\log_b [P(X)] \tag{1}$$

b: < 1, base

If $b = 2$, the unit is the shannon (symbol Sh), often called a 'bit'; if $b = e$, the unit is the natural unit of information (nat); $b = 10$, the hartley (Hart).

Self-information: Definition (II)

Proof (Axiom 3).

$$I(XY) = -\log[P(XY)] = -\log[P(X)P(Y)] = -\log[P(X)] - \log[P(Y)] = I(X) + I(Y) \quad (2)$$

□

Definition (the self-information of a random variable)

For a random variable X with the probability measure P, the self-information is a function of the random variable,

$$I_X(x) := -\log_b[P(X = x)] \quad (3)$$

Entropy

What is the average level of information needed to describe a random variable?

Definition (entropy)

For a discrete random variable X with probability distribution $p : \mathcal{X} \mapsto [0, 1]$, the entropy is

$$H(X) := \mathbb{E}(I(x)) = -\sum_{x \in \mathcal{X}} p(x) \log [p(x)] \tag{4}$$

What is the average level of information that a random variable can offer?

Conditional Entropy

What is the average level of information needed to describe one random variable if another random variable is known?

Definition (conditional entropy)

$$H(Y|X) = -\sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x,y)\, p(y|x) \tag{5}$$

Relative Entropy

On average, how much information do you need to realize that you previously misidentified one random variable P as another Q?

Definition (relative entropy; Kullback–Leibler divergence)

For discrete probability distributions P and Q defined on the same sample space $\mathcal{X}$, the relative entropy from Q to P is

$$D_{KL}(P\|Q) = -\sum_{x\in\mathcal{X}} P(x)\log\left(\frac{Q(x)}{P(x)}\right) \tag{6}$$

On average, how much information do you gain if you are told that you previously misidentified one random variable P as another Q?

Mutual-information

How much information can we obtain about one random variable by observing the other random variable?

Definition (mutual-information)

TODO

References

References i

[1] Wikipedia contributors, *Information content — Wikipedia, the free encyclopedia*, https://en.wikipedia.org/w/index.php?title=Information_content&oldid=1223826857, [Online; accessed 21-May-2024], 2024.

[2] Wikipedia contributors, *Mutual information — Wikipedia, the free encyclopedia*, https://en.wikipedia.org/w/index.php?title=Mutual_information&oldid=1224736345, [Online; accessed 21-May-2024], 2024.

[3] Wikipedia contributors, *Kullback-leibler divergence — Wikipedia, the free encyclopedia*, https://en.wikipedia.org/w/index.php?title=Kullback%E2%80%93Leibler_divergence&oldid=1224220926, [Online; accessed 21-May-2024], 2024.