

Math in tidyinftheo

Andy Pohl

2017-11-29

This package just adds some tidyverse-style enhancements to similar routines in the `infotheo` package (Meyer 2014).

Math

Suppose we have $f(x)$ as a special case of $\log_2(x)$:

$$f(x) = \begin{cases} \log_2(x) & \text{if } x > 0 \\ 0 & \text{if } x \text{ is } 0 \end{cases}$$

We compute Shannon Entropy $H(X)$, as:

$$H(X) = - \sum_{x \in X} p(x) f(x)$$

with the function `shannon_entropy(.data, ..., na.rm=FALSE)`. Conditional Shannon Entropy $H(X|Y)$ as:

$$H(X|Y) = - \sum_{y \in Y} \sum_{x \in X} p(x|y) f(p(x|y))$$

with the `shannon_cond_entropy(.data, ..., na.rm=FALSE)` function. These two entropy equations are enough for the equation for Mutual Information $MI(X; Y)$:

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

and the normalized version of that:

$$NMI(X; Y) = \frac{2 \times MI(X; Y)}{H(X) + H(Y)}$$

using the `mutual_info(.data, ..., normalized=FALSE, na.rm=FALSE)` function. See *Elements of Information Theory* (Cover and Thomas 2001) for a thorough explanation of the mathematics. Also, see the `infotheo` package (Meyer 2014) if additional measures or functionality is needed.

References

Cover, Thomas M., and Joy A. Thomas. 2001. *Elements of Information Theory*. 2nd ed. 10th Ser. New York, NY: John Wiley & Sons, Inc.

Meyer, Patrick E. 2014. *Infotheo: Information-Theoretic Measures*. <https://CRAN.R-project.org/package=infotheo>.