probability distribution

Authors

Alexander Jung*, Department of Computer Science, Aalto University, Espoo, Finland
*Corresponding author: alex.jung@aalto.fi

Abstract

Interpreting data points as independent and identically distributed (i.i.d.) realizations of a random variable (RV) is a basic device for the analysis of machine learning (ML) methods. The typical properties of such data points are governed by the probability distribution of this RV. The probability distribution of a binary RV is fully specified by a single probability. The probability distribution of a real-valued RV might be specified by a probability density function (pdf). In the most general case, a probability distribution is defined by a probability measure.

Definition

To analyze machine learning (ML) methods, it can be useful to interpret data points as independent and identically distributed (i.i.d.) realizations of a random variable (RV). The typical properties of such data points are then governed by the probability distribution of this RV. The probability distribution of a binary RV $y \in \{0,1\}$ is fully specified by the probabilities $\mathbb{P}(y=0)$ and $\mathbb{P}(y=1) = 1-\mathbb{P}(y=0)$. The probability distribution of a real-valued RV $x \in \mathbb{R}$ might be specified by a probability density function (pdf) $p(x)$ such that $\mathbb{P}(x \in [a,b]) \approx p(a)|b-a|$. In the most general case, a probability distribution is defined by a probability measure (Billingsley, 1986; Gray, 2009).

Cross References

i.i.d., realization, RV, probability, pdf

References

1. Billingsley (1986). Probability and Measure. 2nd ed., Wiley, New York, NY, USA.

2. Gray (2009). Probability, Random Processes, and Ergodic Properties. 2nd ed., Springer Science+Business Media, New York, NY, USA.