Abstract

Running inference algorithms on a huge quantity of data knows some perturbations and looses performance. One of Big Data aims is the design of fast inference algorithms able to extract hidden information on a big quantity of data. This paper proposes a new low complexity algorithm for probability density estimation given partial observations. In order to reduce the complexity of the algorithm, a finite numerical data support is adopted in this work and observations are classified by frequencies to reduce their number without loosing significance. By frequency classification we mean, the mapping from the space containing all observed values to a space containing each observable value associated with its observation frequency. This approach relies on Lagrange interpolation for approximating the frequencies with a polynomial function and then build the probability density function. To prove the reliability of
the approach, a simulation is done and results shows the convergence of discussed parameters to the expected values. Big Data field can benefit considerably from proposed approach to achieve density estimation algorithms goal with low cost.

References


Index Terms

Computer Science

Algorithms

Keywords

Probability Density Estimation  Big Data  Polynomial Interpolation  Classification  Algorithms and Complexity