

DATA MINING

wikipedia.com -

data mining is the process of extracting hidden patterns from large amounts of data

equivalently to machine learning, it provides a framework concentrating traditional-statistics power into a well-delimited field of application [data forecasting]

main differences between both theories

- data mining has not accomplished the intended objective [becoming a focused-enough theory] as well as machine learning has done
- data mining does not rely exclusively on statistics, but on further non-strictly-numerical methodologies [lack of pure mathematics essence and thus no direct trendingBot competitors]

an important concept for this theory is “knowledge discovery in databases” (KDD) - the process of finding interesting, interpreted, useful and novel data - thus, data mining is just a part of this more-global idea

processes of KDD [wikipedia]

there are some steps common to all the methods

pre-processing

transform raw data into suitable-for-the-given-method inputs [e.g., independents vs. dependent]

data mining

tasks being commonly carried out by the different methods

1. classification – common algorithms: nearest neighbour, naive bayes classifier or neural networks
2. clustering – second-stage & non-predefined (→ contrariwise task 1.) “classification”
3. regression – genetic programme [not statistical regressions, but decision-trees-based equations]
4. association rule learning – fully-non-numerical [= logical/intuitional] associations among variables

interpreting the results

the necessary common-sense based post-step correcting any wrong-understanding, derived from any algorithm misconception [e.g., avoiding over-fitting]