
Input (X)
Program
Output (y)


Input (X)
Output (y)
Machine learning
Representative Model of the program

Machine Learning

- Learns From Experience
- Experience comes in the form of data
- Data that is used to teach the machines is called training data.

WHAT IS DATA SCIENCE?

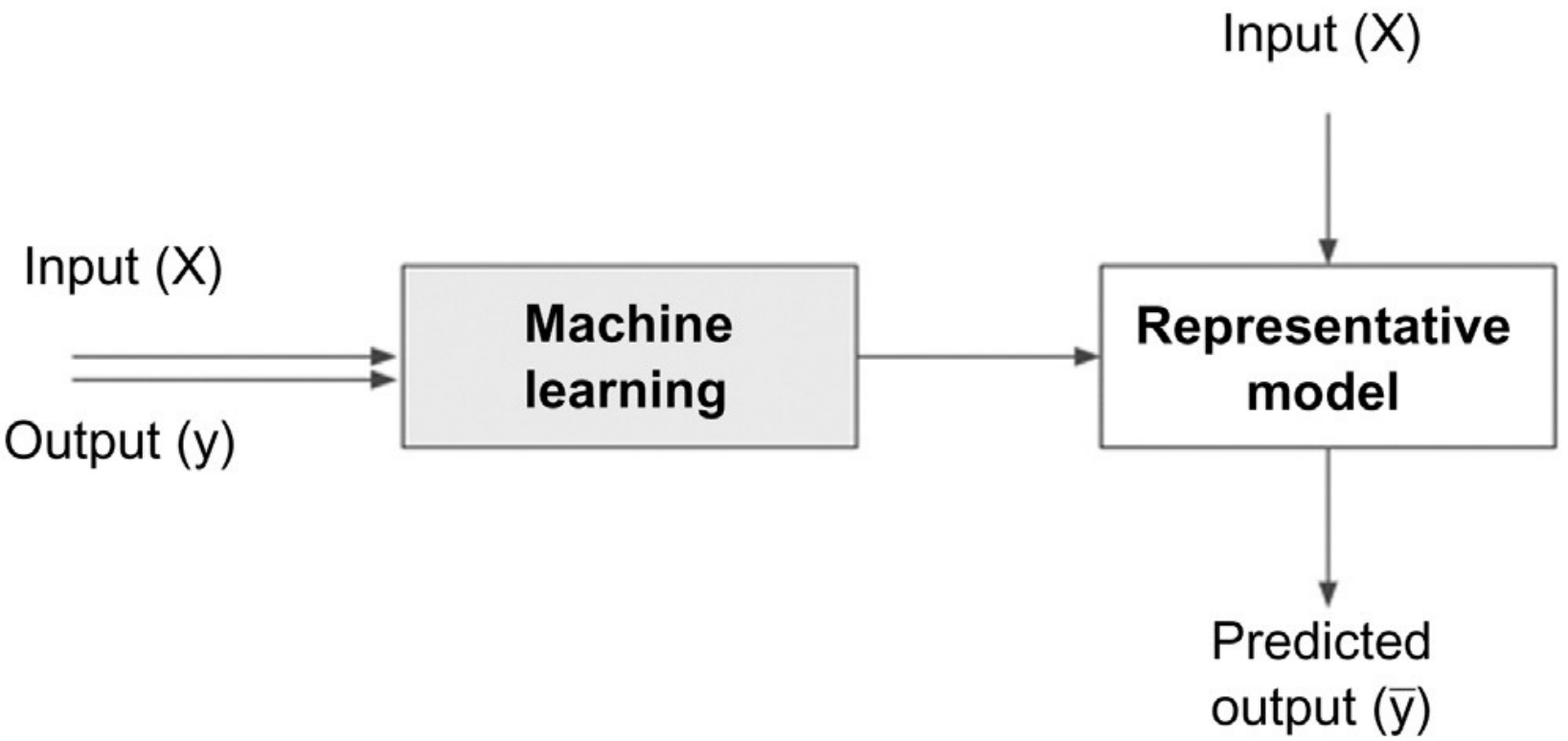
- Data science utilizes certain specialized computational
- methods in order to discover meaningful and useful structures within a dataset.

Extracting Meaningful Patterns

- Knowledge discovery in databases is the nontrivial process of identifyingunderstandable patterns or relationships within a dataset in order to make important decisions
- One of the key aspects of data science is the process of generalization of patterns from a dataset
- The generalization should be valid, not just for the dataset used to observe the pattern, but also for new unseen data.

Building Representative Models

- Modeling is a process in which a representative abstraction is built from the observed dataset.
- Data science is the process of building a representative model that fits the n observational data.
- A Model can be used for both predictive and explanatory applications.


Input (X)
Input (X)
Output (y)
Machine learning
Representative model
Predicted output (ȳ)

Combination of Statistics, Machine Learning, and Computing

- The algorithms used in data science originate from these disciplines
- One of the key ingredients of successful data science is substantial prior knowledge about the data and the business processes that generate the data, known as subject matter expertise.
- Like many quantitative frameworks, data science is an iterative process in which the practitioner gains more information about the patterns and relationships from data in each cycle

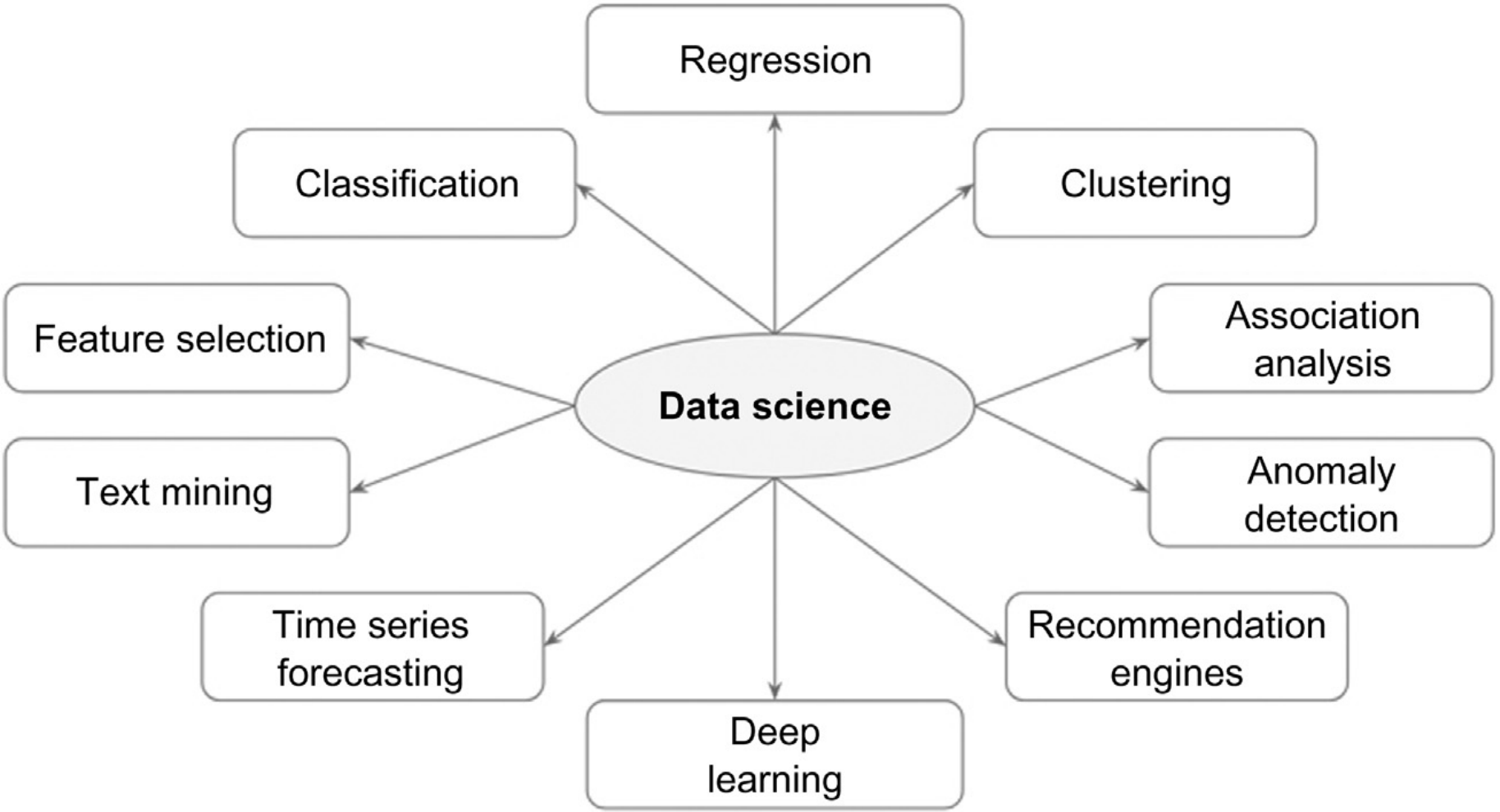
Learning Algorithms

- Data science is also defined as a process of discovering previously unknown patterns in data using automatic iterative methods.
- The application of sophisticated learning algorithms for extracting useful patterns from data differentiates data science from traditional data analysis techniques.
- Based on the problem, data science is classified into tasks such as classification, association analysis, clustering, and regression.

Associated Fields

- Descriptive statistics
- Exploratory visualization
- Dimensional slicing
- Hypothesis testing
- Data engineering
- Business intelligence

DATA SCIENCE CLASSIFICATION


Regression
Classification
Clustering
Feature selection
Association analysis
Data science
Text mining
Anomaly detection
Time series forecasting
Recommendation engines
Deep learning