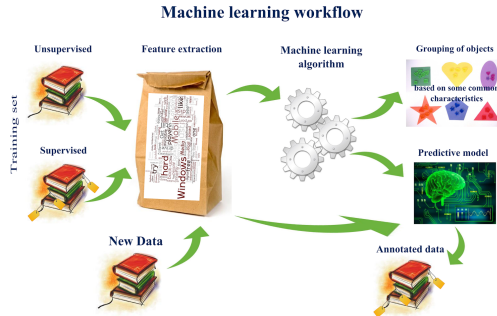


Agenda

- 1 What is the Machine learning?
- 2 Types of machine learning.
 - 1 Supervised learning.
 - 2 Unsupervised learning.
 - 3 Semi-supervised learning.
 - 4 Reinforcement learning.
- 3 Machine learning techniques
 - 1 Classification.
 - 2 Regression.
 - 3 Clustering.

What is machine learning?

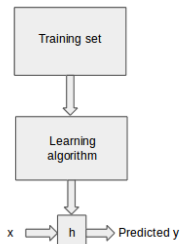
Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.



Supervised learning

Supervised learning (predictive model, "labeled" data) - learning some mapping from input data to output.

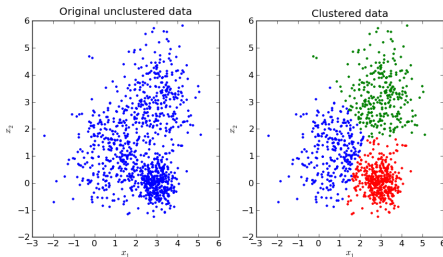
- Classification (Logistic Regression, Decision Tree, KNN, Random Forest, SVM, Naive Bayes, etc)
- Numeric prediction (Linear Regression, KNN, Gradient Boosting & AdaBoost, etc)



Unsupervised learning

Unsupervised learning (descriptive model, "unlabeled" data) - finds hidden data structure from unlabeled data.

- clustering (K-Means, Gaussian mixtures, DBSCAN and etc.)
- pattern discovery
- dimension reduction (PCA, linear discriminant analysis)



Semi-supervised learning

Semi-supervised learning is a class of supervised learning tasks and techniques that also make use of unlabeled data for training – typically a small amount of labeled data with a large amount of unlabeled data.

- unlabeled data is cheap
- labeled data can be hard to get
- human annotation is boring
- labels may require experts

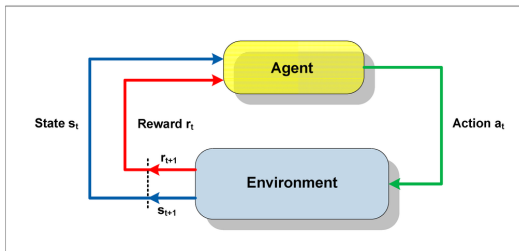
Semi-supervised algorithms

- Self Training
- Generative Models
- S3VMs
- Graph-Based Algorithms
- Multiview Algorithms

Reinforcement learning

Reinforcement learning is learning what to do—how to map situations to actions—so as to maximize a numerical reward signal

- Rules are unknown
- No supervisor, only reward signal
- Agent's actions affect the subsequent data it receives



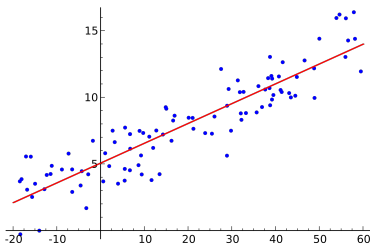
Classification

- Data is labeled. It means it is assigned a class, for example spam/non-spam or fraud/non-fraud.
- The decision being modelled is to assign labels to new unlabelled pieces of data.
- This can be thought of as a discrimination problem, modelling the differences or similarities between groups.
- Usually supervised



Regression

- Data is labelled with a real value (think floating point) rather than a label.
- Examples that are easy to understand are time series data like the price of a stock over time.
- The decision being modelled is what value to predict for new unpredicted data.



Clustering

Data is not labeled, but can be divided into groups based on similarity and other measures of natural structure in the data.

Algorithms

- Connectivity-based clustering (hierarchical clustering)
- Centroid-based clustering (k-means clustering)
- Density-based clustering (DBSCAN)

