Machine Learning Approaches. Overview

Alexander Trofimov
PhD, professor, NRNU MEPhI

lab@neuroinfo.ru
http://datalearning.ru

Course "Machine Learning"

November 2022

What is Machine Learning?

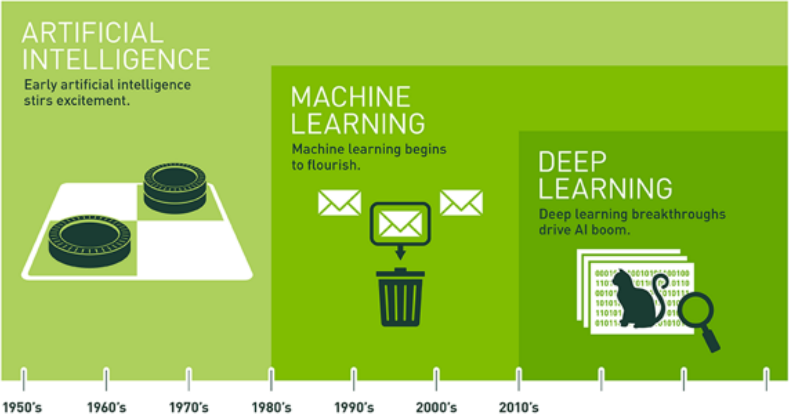
Machine learning is the science (and art) of programming computers so they can learn from data

> "Machine learning is field of study that gives computers the ability to learn without being explicitly programmed."
>
> — Arthur Samuel, in IBM Journal of Research and Development, 1959

The term "Machine learning" is related to:

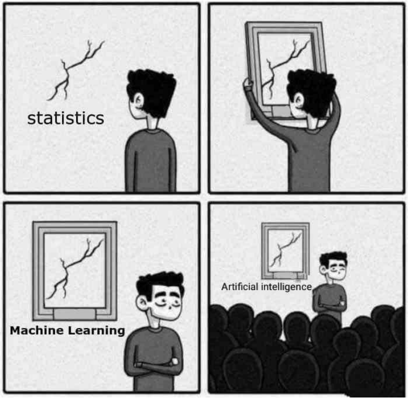
- Artificial intelligence
- Mathematical statistics
- Data science

Machine Learning and Artificial Intelligence



When we talk about AI we primarily refer to two specific areas: machine learning and deep learning

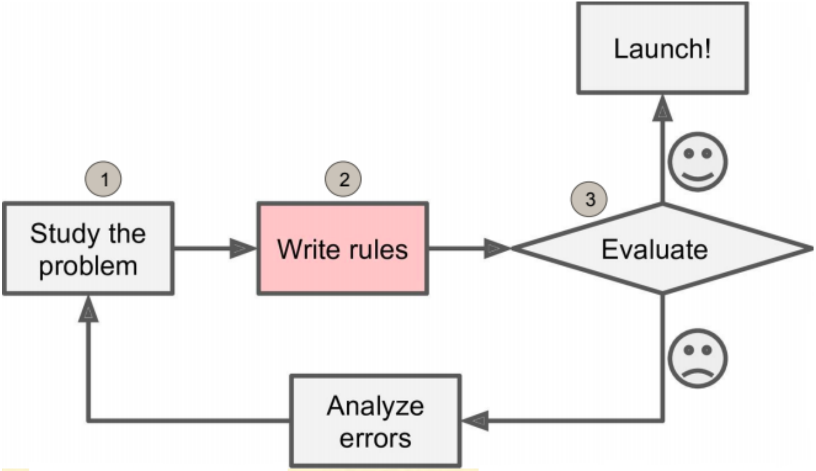
Machine Learning and Mathematical Statistics



"When you're fundraising, it's AI. When you're hiring, it's ML. When you're implementing, it's logistic regression."

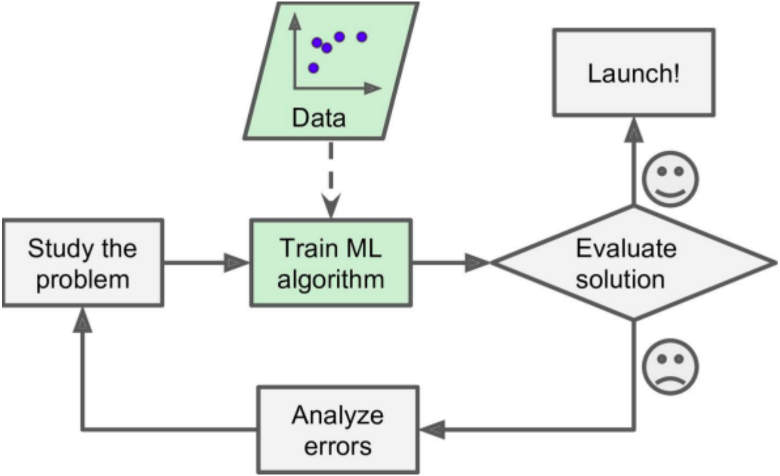
— everyone on Twitter ever

Traditional Approach



Since the problem is difficult, your program will likely become a long list of complex rules that pretty hard to maintain

Machine Learning Approach



Machine learning approach is useful for problems that are too complex for traditional approaches or have no known algorithms

Traditional Programming vs Machine Learning

Traditional programming:

Machine learning:

Approaches to Modelling

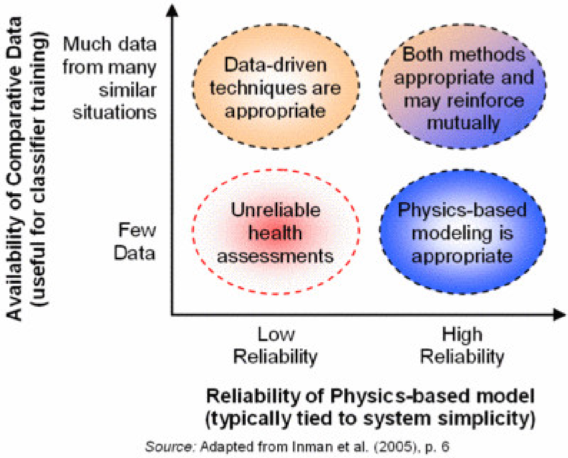
Approaches to modelling:

- Model-based
  Use known physical, economical, biological, etc. laws and expert rules to construct the model
- Data-driven
  Use data to construct the model

Key features of data-driven approach:

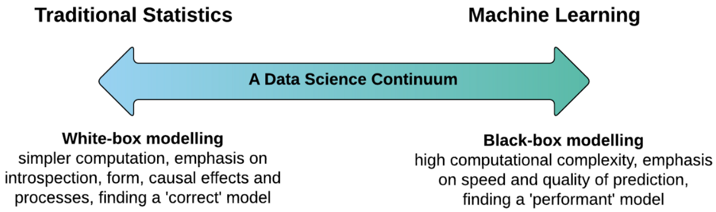
- The precise mathematical model is absent or unacceptably complex
- There are statistical "input-output" data about the system under modelling
- The data processing algorithm is unknown a priori, it is a result of learning procedure

Data-driven vs Model-based Approaches



The model-based approach is generally more robust in the sense that it can deal more easily with new or unforeseen situations

Traditional Statistics vs Machine Learning

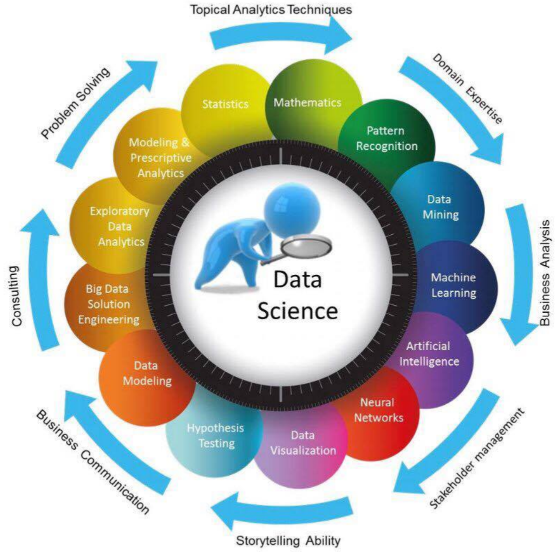


Most machine learning systems require the ability to explain why certain predictions are made

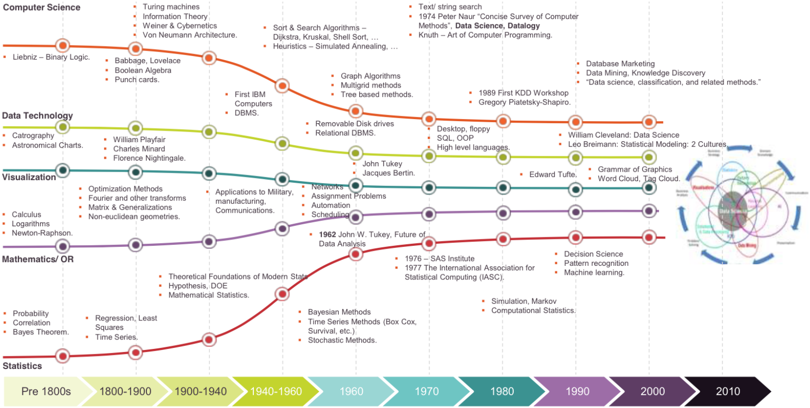
- **Black-box models:** accurate but difficult to explain
  Neural networks, complicated ensembles, etc.
- **White-box models:** weaker but usually simple to explain
  Linear regression, decision trees, etc.

We define explainability-accuracy trade-off when choosing a machine learning model

Machine Learning and Data Science



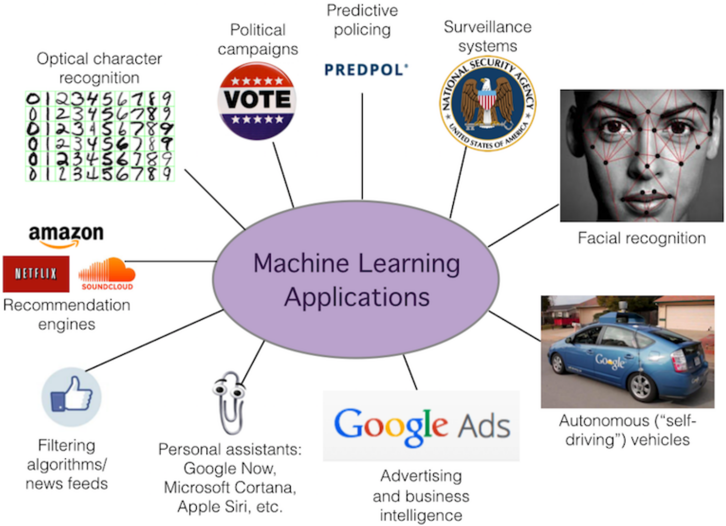
History of Data Science



Machine Learning Milestones

1997 IBM's Deep Blue beats the world champion at chess
2005 Autonomous ground vehicles: DARPA Grand Challenge
2006 Google Translate
2011 DARPA CALO project, Apple Siri
2011 IBM's Watson beats two human champions in a Jeopardy! competition
2012 The Google Brain team create a neural network that learns to recognize cats by watching unlabeled images taken from frames of YouTube videos
2014 Facebook DeepFace identifies faces with 97% accuracy
2015 OpenAI by Elon Musk and Sam Altman - $1 bln.
2016 OpenAI, Google's DeepMind: Atari games
2016 Google's AlphaGo beats the world champion at Go
2018 Tesla launches self-driving vehicle
2020 Baidu Inc. launches robotaxi service Apollo Go
2020 DeepMind AlphaFold – predicting the results of protein folding

Machine Learning Applications



Machine Learning Competitions

http://www.netflixprize.com/ http://www.kaggle.com/ http://dataring.ru/

	Passenger Screening Algorithm Challenge Improve the accuracy of the Department of Homeland Security's threat recognition algorithms Featured · 3 months to go	**$1,500,000** 239 teams
Zillow	**Zillow Prize: Zillow's Home Value Prediction (Zestimate)** Can you improve the algorithm that changed the world of real estate? Featured · 4 months to go	**$1,200,000** 2,638 teams
	Carvana Image Masking Challenge Automatically identify the boundaries of the car in an image Featured · 22 days to go	**$25,000** 546 teams
En	**Text Normalization Challenge - English Language** Convert English text from written expressions into spoken forms Research · 2 months to go	**$25,000** 5 teams
Ru	**Text Normalization Challenge - Russian Language** Convert Russian text from written expressions into spoken forms Research · 2 months to go	**$25,000** 3 teams
	Web Traffic Time Series Forecasting Forecast future traffic to Wikipedia pages Research · 7 days to go	**$25,000** 1,020 teams
	Personalized Medicine: Redefining Cancer Treatment Predict the effect of Genetic Variants to enable Personalized Medicine Research · a month to go	**$15,000** 1,066 teams

CRISP-DM

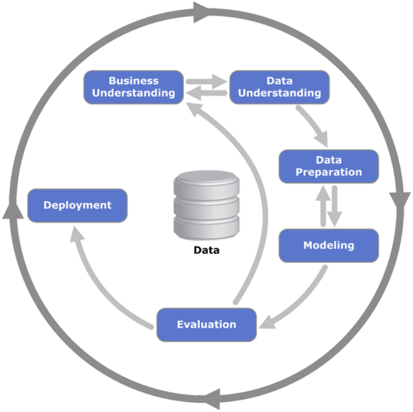
Cross-Industry Standard Process for Data Mining (CRISP-DM) is the commonly used data mining methodology (1999)

Companies:

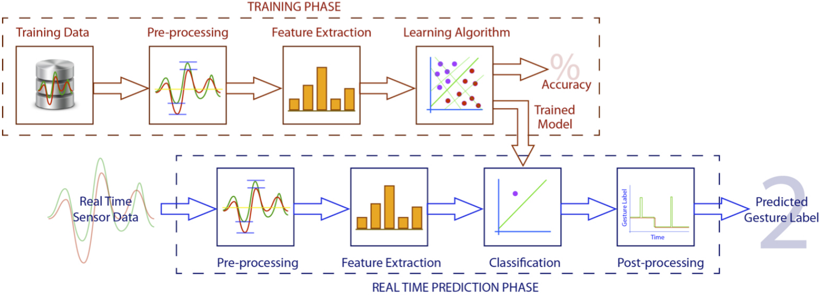
- SPSS
- Teradata
- Daimler AG
- NCR Corporation
- OHRA

IBM has released a new methodology for Data Mining/Predictive Analytics projects in 2015 called Analytics Solutions Unified Method (ASUM-DM) which refines and extends CRISP-DM

Data mining life cycle:



Machine Learning Pipeline



- Training phase
- Prediction phase

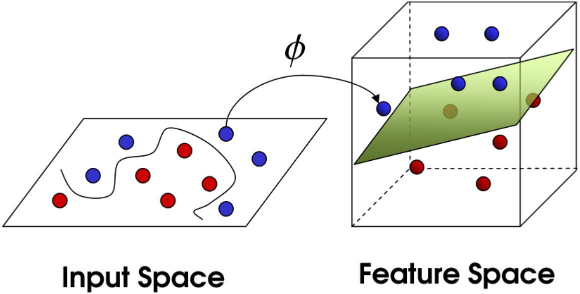
- Pre-processing
- Feature extraction
- Learning algorithm
- Accuracy estimation
- Post-processing

Feature Engineering

Feature engineering asks: **what is the best representation of the sample data to learn a solution to your problem?**

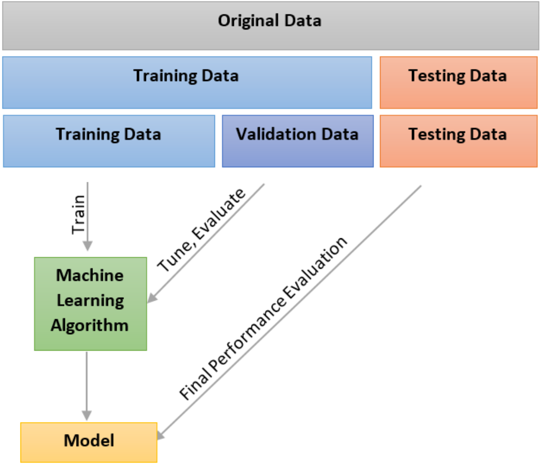
Better features mean:

- Flexibility
- Simpler models
- Better results



Accuracy Estimation

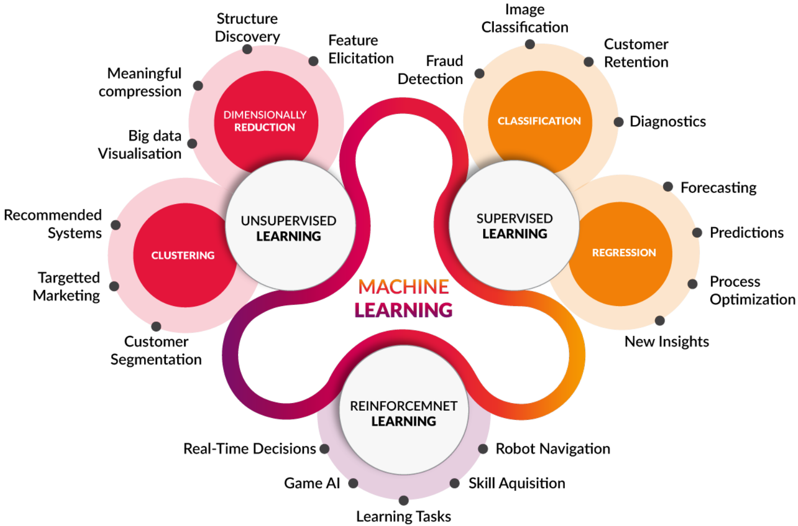
Model estimation asks: **how can we get an unbiased estimate of the accuracy of a learned model?**



Machine Learning Paradigms

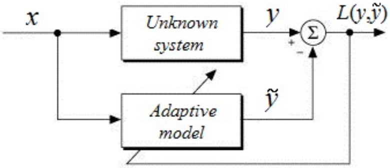
- Supervised learning
- Unsupervised learning
- Reinforcement learning
- Ensemble learning
- Semi-supervised learning
- Active learning
- Deep learning
- Transfer learning
- ...

Machine Learning Tasks



Supervised Learning. Problem Statement

Given:
$\mathscr{D} = \{(x^{(1)}, y^{(1)}), ..., (x^{(n)}, y^{(n)})\}$
$x^{(i)} \in \mathscr{X}, \quad i = \overline{1, n}$
$y^{(i)} \in \mathscr{Y}, \quad i = \overline{1, n}$



Criterion:

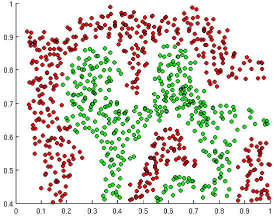
$$R^* = \frac{1}{n} \sum_{i=1}^{n} Loss\left(y^{(i)}, \tilde{y}^{(i)}\right) \to \min_{w}$$

$\tilde{y}^{(i)}$ is the model's output for input $x^{(i)}, i = \overline{1, n}$
w is a vector of model's parameters
$Loss$ is a loss function

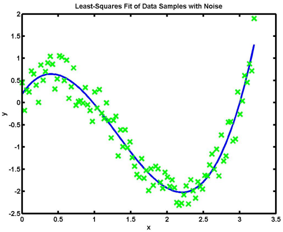
Find out:
Vector w^* that minimizes R^*

Supervised Learning. Types of problems

- Classification
  $\mathscr{D} = \{(x^{(1)}, y^{(1)}), ..., (x^{(n)}, y^{(n)})\}$
  $x^{(i)} \in \mathscr{X}$ is i-th object, $i = \overline{1, n}$
  $y^{(i)} \in \mathscr{Y}$ is label of $x^{(i)}$
  $\mathscr{Y} = \{1, ..., K\}$ is a set of class labels



- Regression
  $\mathscr{D} = \{(x^{(1)}, y^{(1)}), ..., (x^{(n)}, y^{(n)})\}$
  $x^{(i)} \in \mathscr{X}$ is i-th object, $i = \overline{1, n}$
  $y^{(i)} \in \mathscr{Y}$ is response for $x^{(i)}$
  $\mathscr{Y} = \mathbb{R}^L$ is a set of responses



Supervised Learning. Approaches

- Linear regression
- Logistic regression
- Bayesian methods
- Nearest neighbor methods
- Support vector machines
- Decision tree
- Neural networks
- ...

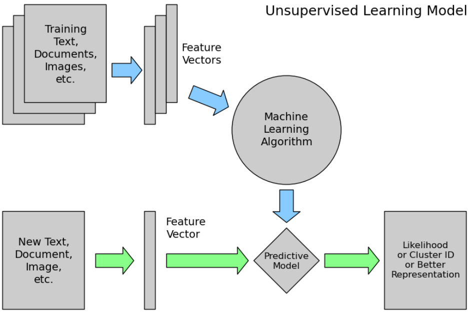
Unsupervised Learning. Problem Statement

Given:

$$\mathscr{D} = \{x^{(1)}, ..., x^{(n)}\}, \quad x^{(i)} \in \mathscr{X}, \quad i = \overline{1, n}$$

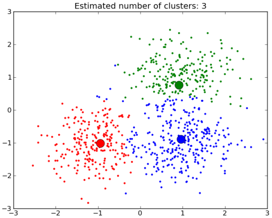
Find out:

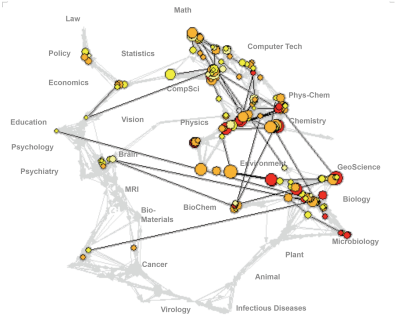
Better representation of $\mathscr{D}$, estimate distribution, detect anomalies, find out patterns etc.

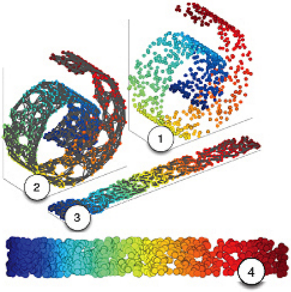


Unsupervised Learning. Types of Problems

- Clustering
- Missing data recovery
- Dimensionality reduction
- Visualization
- Anomaly detection







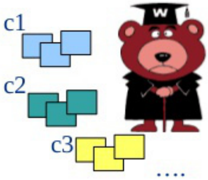
Unsupervised Learning. Approaches

- Cluster analysis
- Self-organizing maps (SOM)
- Principal component analysis (PCA)
- Independent component analysis (ICA)
- Multidimensional scaling (MDS)
- T-distributed Stochastic Neighbour Embedding (t-SNE)
- ...

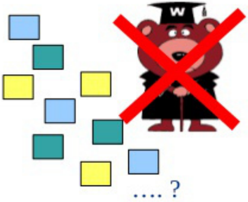
Supervised vs Unsupervised Learning

Supervised Vs. **Unsupervised**

- **Supervised**
  - **knowledge of output -** learning with the presence of an "expert" / teacher
    - data is **labelled** with a class or value
    - **Goal:** predict class or value label
      - e.g. Neural Network, Support Vector Machines, Decision Trees, Bayesian Classifiers ....



- **Unsupervised**
  - **no knowledge of output** class or value
    - data is **unlabelled** or value un-known
    - **Goal:** determine data patterns/groupings
  - Self-guided learning algorithm
      - (internal self-evaluation against some criteria)
      - e.g. k-means, genetic algorithms, clustering approaches ...



Semi-supervised Learning. Problem Statement

Given:

Labeled sample:

$$\mathscr{D}_L = \{(x^{(1)}, y^{(1)}), ..., (x^{(n_L)}, y^{(n_L)})\}$$

$$x^{(i)} \in \mathscr{X}, \quad y^{(i)} \in \mathscr{Y}, \quad i = \overline{1, n_L}$$

$\mathscr{Y} = \{1, ..., K\}$ is a set of class labels

Unlabeled sample:

$$\mathscr{D}_U = \{x^{(n_L+1)}, ..., x^{(n_L+n_U)}\}, \quad n_L \ll n_U$$

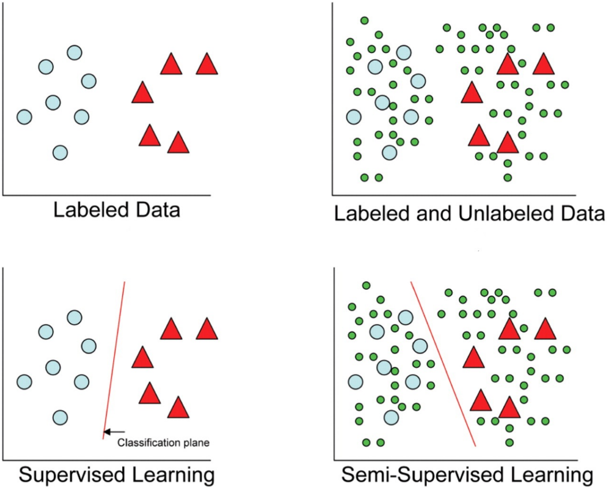
$$x^{(n_L+i)} \in \mathscr{X}, \quad i = \overline{1, n_U}$$

Objectives:

- Construct classification algorithm
- Predict labels for cases from $\mathscr{D}_U$ (transductive learning)

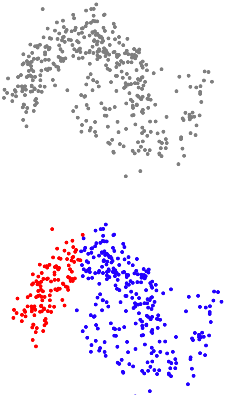
Semi-supervised Learning. Illustration 1

Semi-supervised Learning $\neq$ Supervised Learning



Semi-supervised Learning. Illustration 2

Semi-supervised Learning $\neq$ Unsupervised Learning



Unsupervised clustering

Semi-supervised clustering

Ensemble Learning

Given:

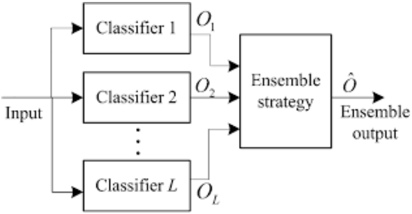
$\mathscr{D} = \{(x^{(1)}, y^{(1)}), ..., (x^{(n)}, y^{(n)})\}$

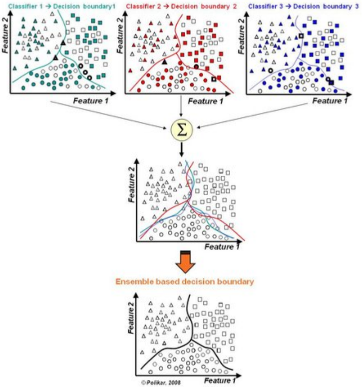
$x^{(i)} \in \mathscr{X}, \quad i = \overline{1, n}$

$y^{(i)} \in \mathscr{Y} = \{1, ..., K\}$ — class labels

$h_1, ...h_L$ — base classifiers (hypotheses)

$h = h_1 \circ ... \circ h_L$





Ensemble Learning. Approaches

Approaches:

- Bayesian voting
- Manipulating the training sample
  - Cross-validated committees
  - Bagging (Bootstrap aggragation)
  - Boosting
- Manipulating the features
- Manipulating the outputs

Algorithms:

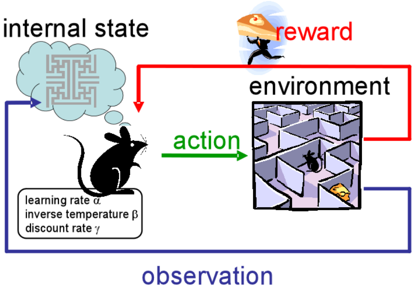
- Voting classifiers and regressors
- AdaBoost
- Random forest
- Extremely randomized trees
- ...

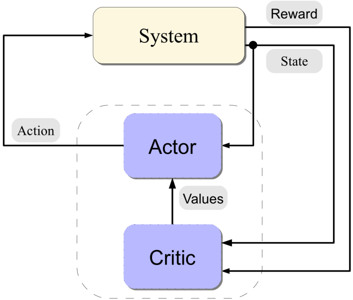
Reinforcement Learning

The model (agent) interacts with its environment in discrete time steps

The agent's objective is to act in the environment so as to maximize some long-term cumulative reward

The train sample is absent, the learning is in on-line mode, the reactions of environment are used

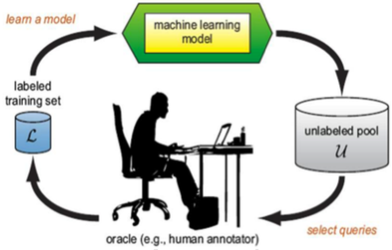


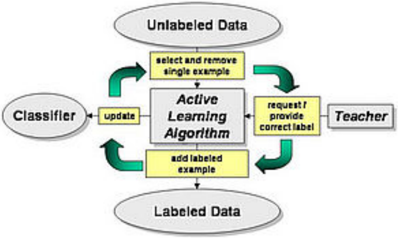


Active Learning

Active learning is a special case of semi-supervised learning in which a learning algorithm is able to interactively query the user (or some other information source) to obtain the desired outputs for given unlabeled data

In statistics literature active learning is known as optimal experimental design

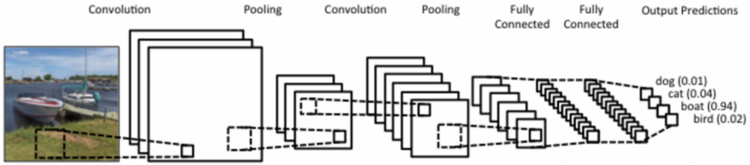




Deep Learning. Approaches

Deep learning (hierarchical learning) is part of a broader family of machine learning methods based on learning data representations, as opposed to task-specific algorithms

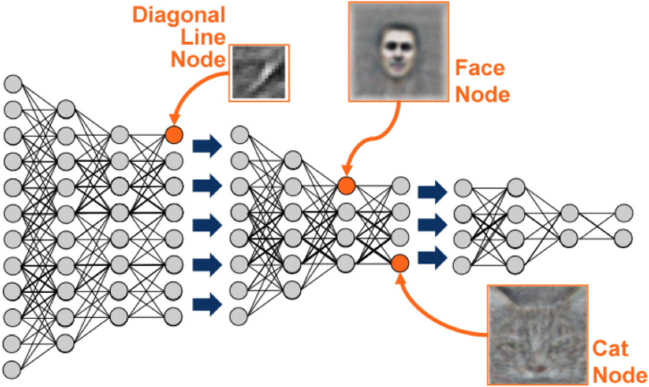
- Convolutional neural networks (CNN)
- Deep belief networks (DBN)
- Deep Boltzmann machine
- Deep recurrent neural network



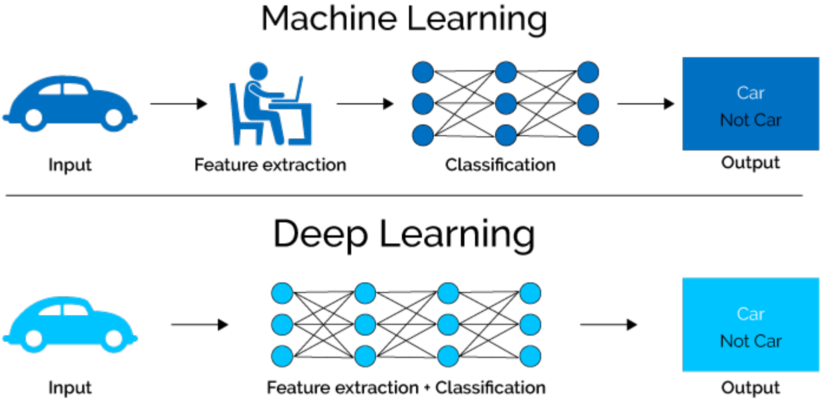
Deep Learning. Hierarchical Feature Representation

Deep Learning

Lots of Data + Neural Nets + Training = Hierarchical & Associational Feature Representation

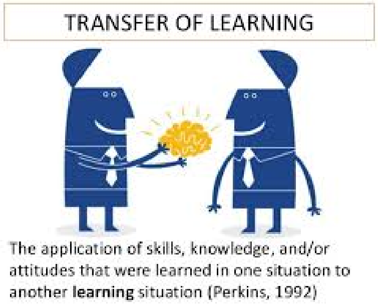


Machine Learning vs Deep Learning



Transfer Learning

Transfer learning (inductive transfer) is a research problem in machine learning that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem



Transfer Learning