Fundamentals of Artificial Neural Networks and Machine Learning

Introduction

- Humans have developed sensory systems to perceive their environments and make decision based on what they observe, for example:

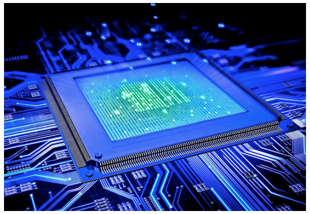
    - Understanding speech
    - Recognizing faces and objects
    - Discerning taste

- This is greatly influenced by how we perceive the patterns in our environment.

Humans have always strived to impart machines with similar capabilities as theirs for faster task execution, better predictability and higher accuracy

Recent Breakthroughs

- IBM's Cat-brain project is a chip like brain (SyNAPSE) to capture surrounding information in real time. Simulated the cat brain cortex with 147,456 cores and 144TB of memory: basic synaptic circuit for the brain chip.



Current work on 10^9 artificial neurons and 10^14 synapses

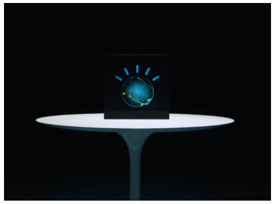
- Google's DeepMind project, aims at formalizing intelligence to implement in machines and to understand the human brain. Uses deep learning technologies.

DEEPMIND

On March 2016, AlphaGo beat the top player in the world (Lee Sedol) in the Go game (Dan 9)

Recent Breakthroughs

- IBM's Watson, QA system is used as knowledge repository, to answer virtually any inquiry (in Wikipedia and other knowledge repositories). Uses NLU and ML technologies. Has access to more than 10^9 pages content.

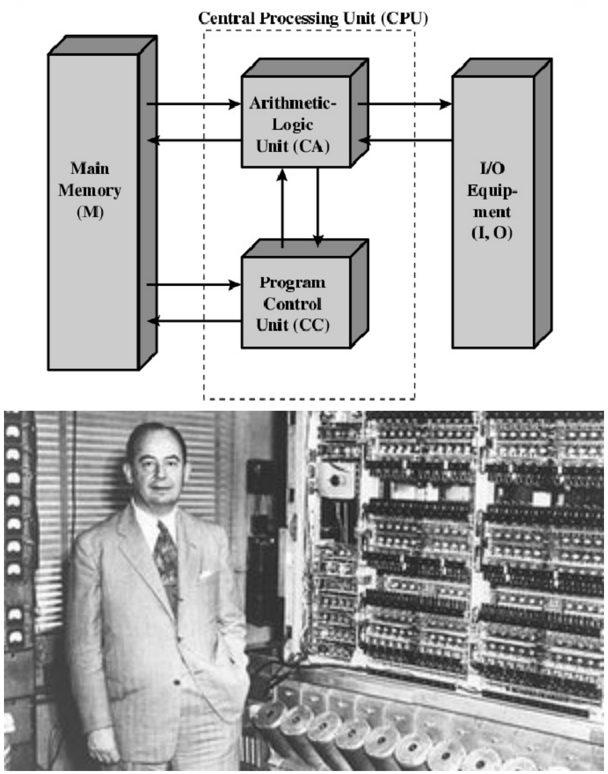


Won Jeopardy! Contest against top players in the world. Project Intu will equip devices and machines with Watson's powerful AI tools, hence providing them with cognitive capabilities and make them aware of their environments (ZDNET, Nov. 2016)

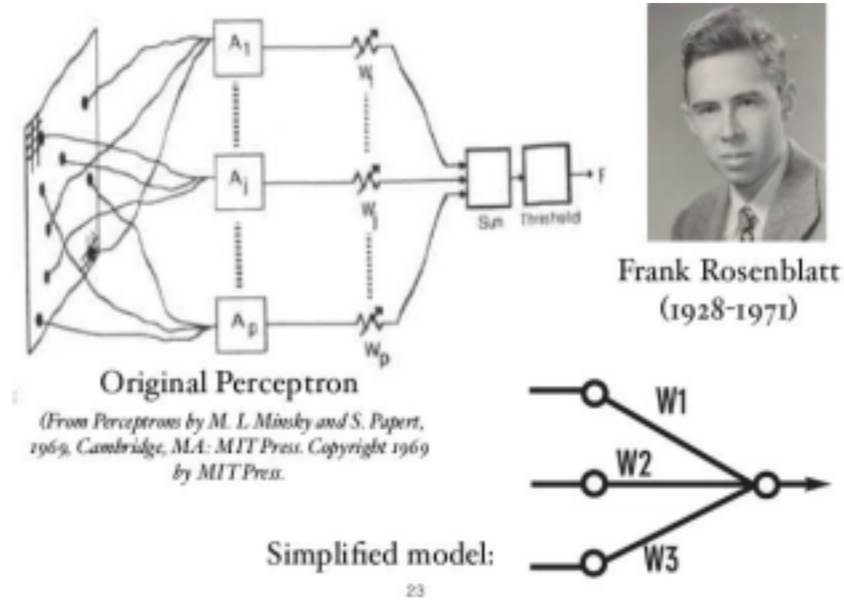
- Amazon, Google and Netflix are using Big Data analytics based on Deep Learning to target consumer behavior in various areas including entertainment, shopping and provide adequate recommendations to the user

Origins of Modern Computing and Machine Learning

Von Neuman Computer



Rosenblatt Perceptron



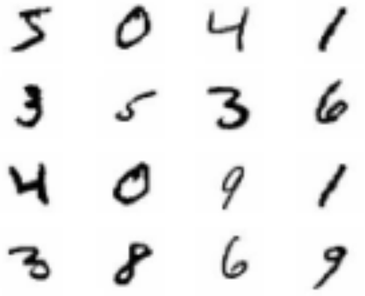
Father of Modern Computing

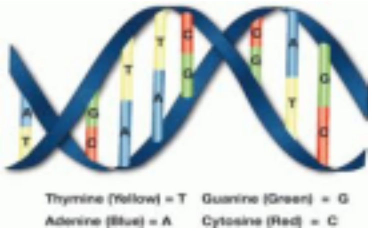
Father of Connectionist Computing

What is a Pattern?

A pattern is an abstract entity that describes a physical object such as:

- Speech signal
- Handwritten words or digits
- Human face
- DNA sequence
- Weather information





What is Pattern Recognition?

- Pattern recognition (PR) is the study of how machines:
  - observe the environment
  - learn to distinguish patterns of interest
  - make correct decisions about the
  - categories of the patterns.

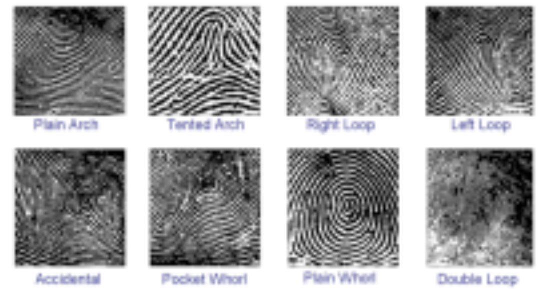
- PR involves theory, algorithms, and systems

  that can assign patterns into categories.

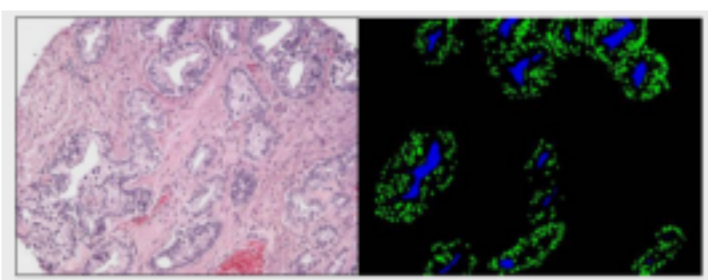
Pattern Recognition Applications

License Plate Recognition

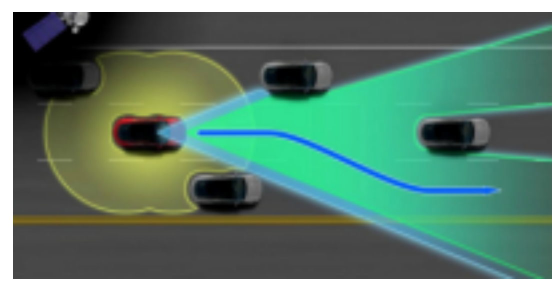
Fingerprint Recognition



Cancer Detection



Autonomous Navigation



Pattern Recognition Algorithms

Major approaches for PR include

- Methods for Regression (Continuous Models):
  - Least Squares
  - Ridge Regression
  - The Lasso Approach

- Method for Classification (Discrete Modeling):
  - Linear and Quadratic Discriminant Analysis
  - Decision Trees
  - Rosenblatt's Perceptron
  - Suupport Vector Machines and ANN

Pattern Recognition Challenges

- MNIST dataset

A zero that's difficult to distinguish from a six algorithmically

- ImageNet challenge database

Pattern Recognition Challenges

Main Challenges of PR:

- Wide variability of the same digits and objects.
- How to differentiate between "6" and "0"? it quickly becomes quite complicated to compile a list of heuristics.
- If we have more than 5 objects? How can we solve this problem?
- Nearly impossible to create handcrafted rules to understand each pixel.

Objective: design a machine that learns patterns from data!

Solution: Machine Learning!!

Machine Learning

- Definition: An area of artificial intelligence (AI) that provides computers with the ability to learn from data while discovering patterns.

- It can be applied in situations where it is very challenging or impossible to define heuristic/manual rules e.g., Face detection, Speech recognition, Stock prediction.

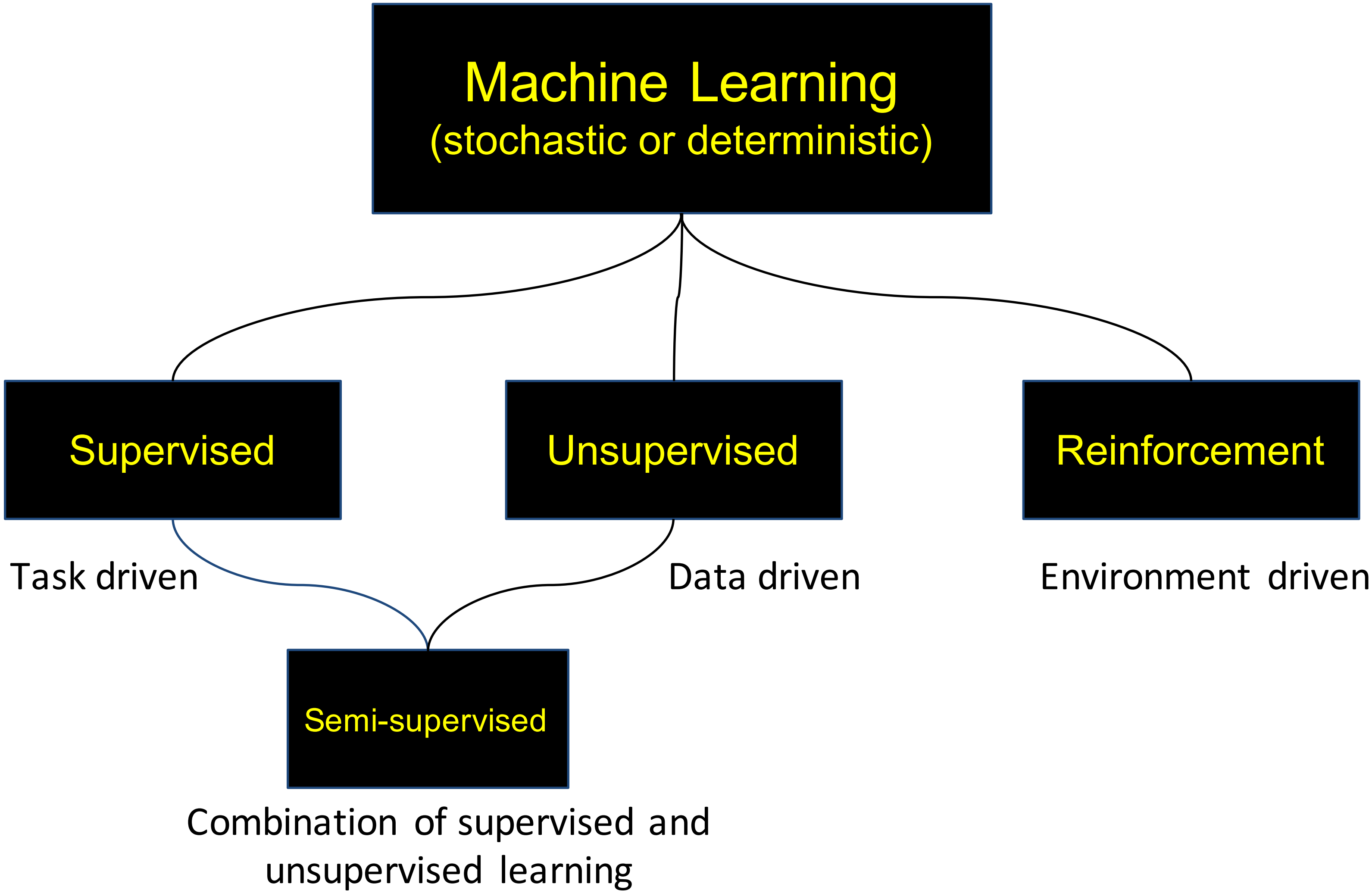
Machine Learning

Formally speaking: A ML algorithm improves some measures of performance *P* when executing some task *T* through some type of experience *E*

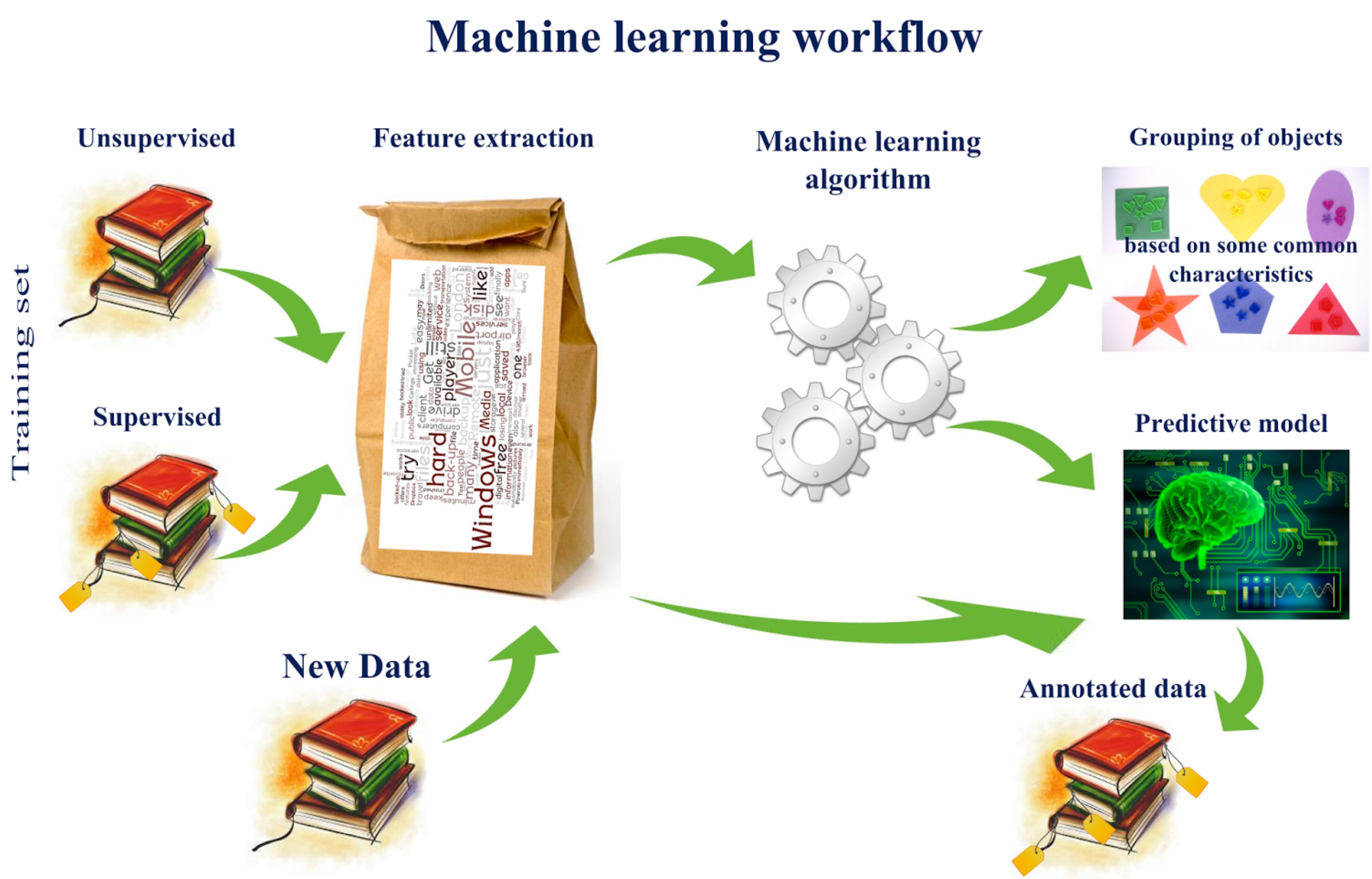
Example: Learning to Detect Credit Card Fraud:

- Task *T: assign category of Fraud/Not Fraud to CC transaction*
- Performance Measure *P: Accuracy of the classifier, i.e, assign higher penalty when Fraud is labeled as Not Fraud*
- Training Process *E: Set of historical credit card training transactions categorized as Fraud or NotFraud*

Major Classes of Machine Learning (ML)

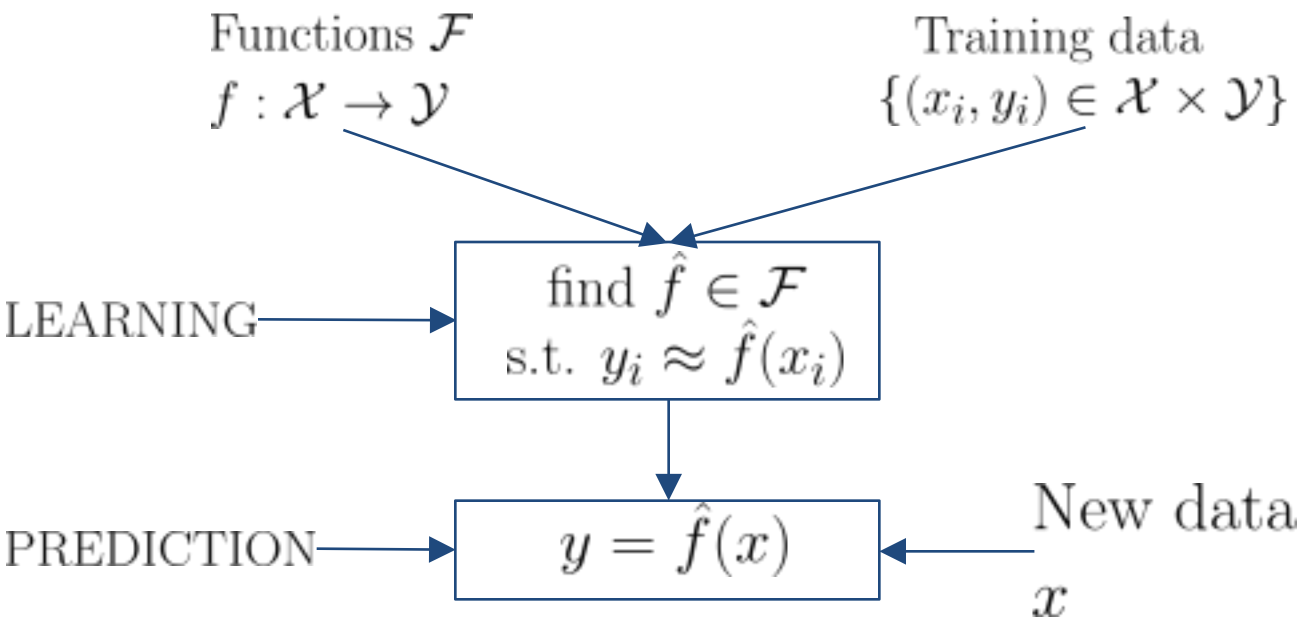


Machine Learning Working Mechanism



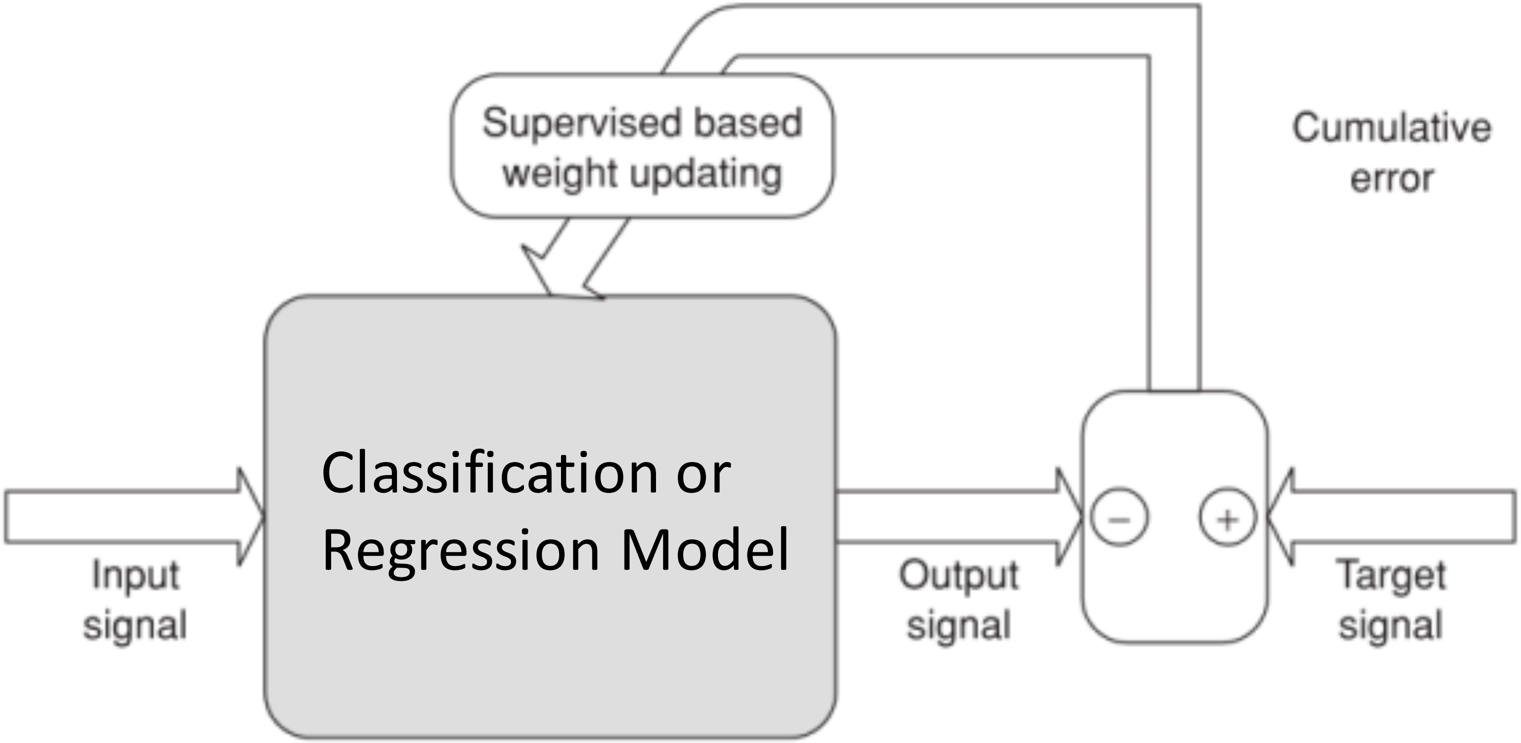
Supervised Learning: Overview

The corresponding classes of the training data are known: "labeled data"



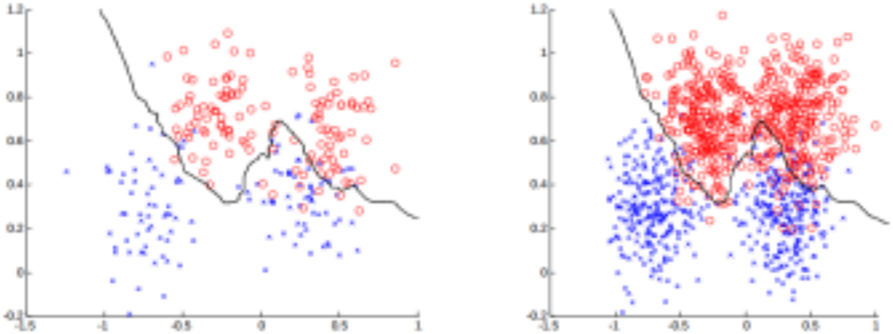
Supervised Learning: Overview

- Generic architecture of supervised learning

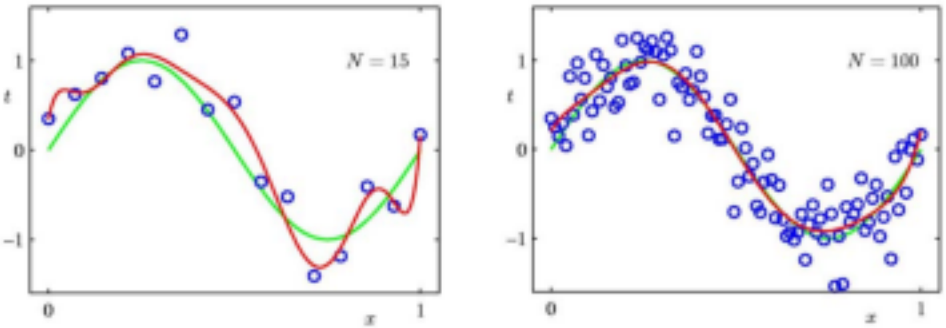


Supervised Learning: Overview

Classification



Regression



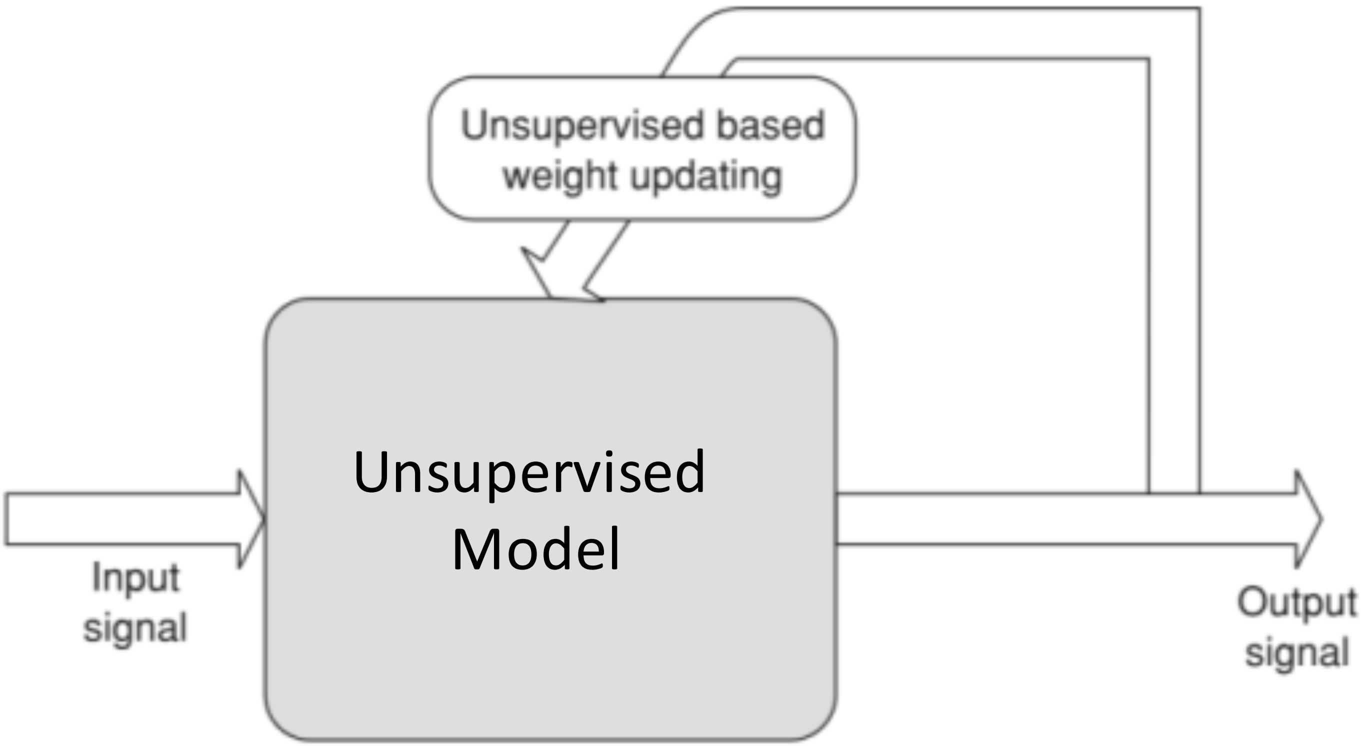
- Objective: to do well on test data that is completely unseen during learning.
- This is done by cross-validation, regularization, dropout, etc.
- Well-known methods: LDA, K-NN, Neural Networks, SVM, etc.

Supervised Learning: Algorithms

Algorithms	Applications
Decision Trees.	Drug analysis, financial analysis, medical text classification.
Naïve Bayes Classification.	Spam filter, news articles classification, face recognition.
Ordinary Least Squares Regression.	Stock prediction, weather prediction, earthquake prediction.
Support Vector Machines.	Human splice site recognition, image-based gender detection, large-scale image classification.
Artificial Neural Networks.	game-playing and decision making, face identification, speech recognition, automated trading systems.

Unsupervised Learning: Overview

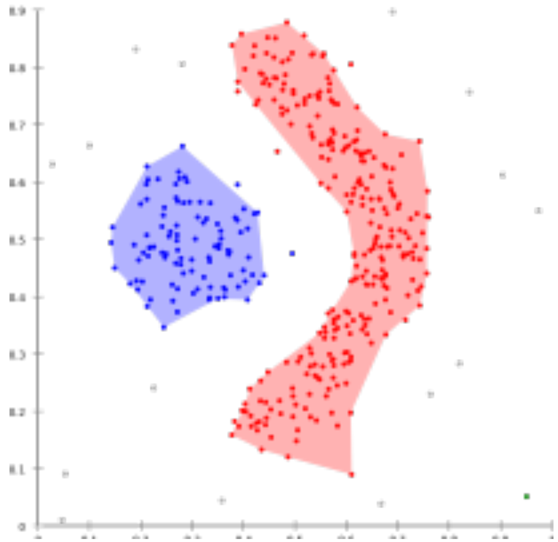
- Generic architecture of unsupervised learning



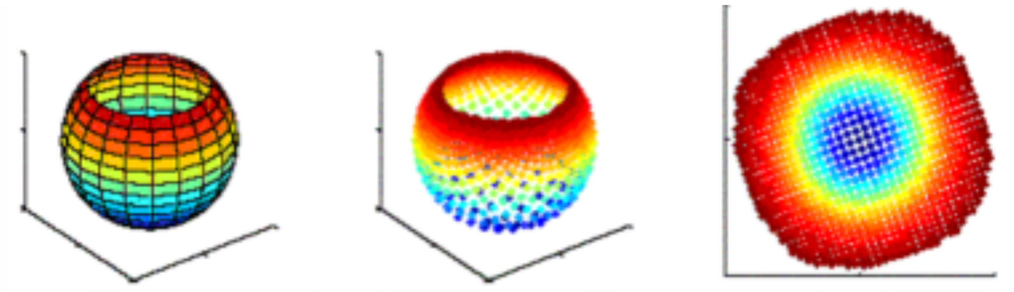
Unsupervised Learning: Overview

- The correct classes of the training data are unknown.

Clustering



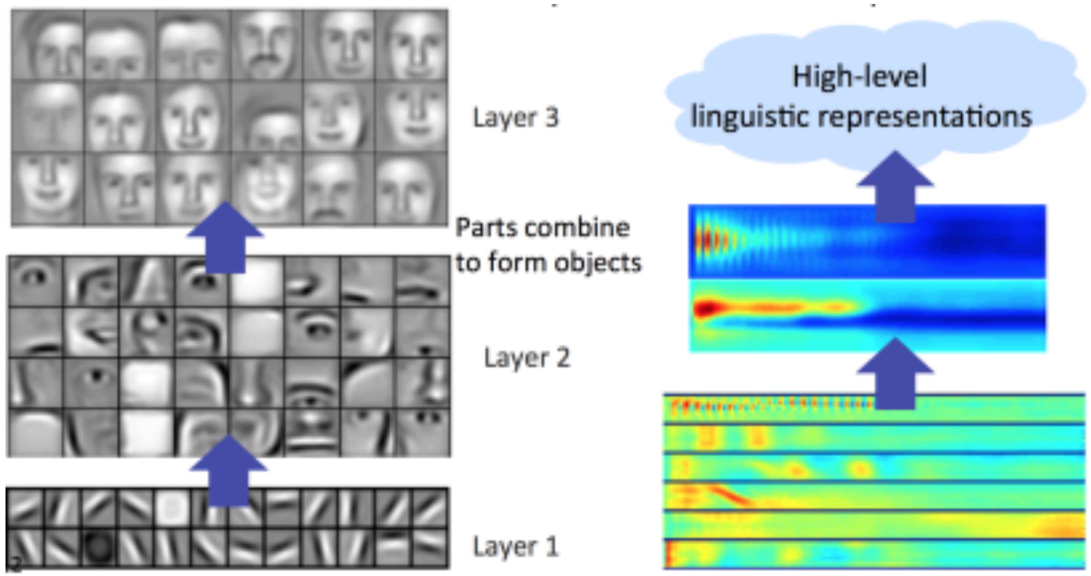
Dimensionality Reduction



- Both can be used to extract/select features.
- Well known methods: Kohonen's SOM, K-means, DBSCAN, PCA, etc.

Modern Unsupervised Learning

- Feature Representation Learning



- Well-known methods: RBMs, Auto-encoders.
- They can be used to learn new feature representations or to initialize neural networks.

Unsupervised Learning: Algorithms

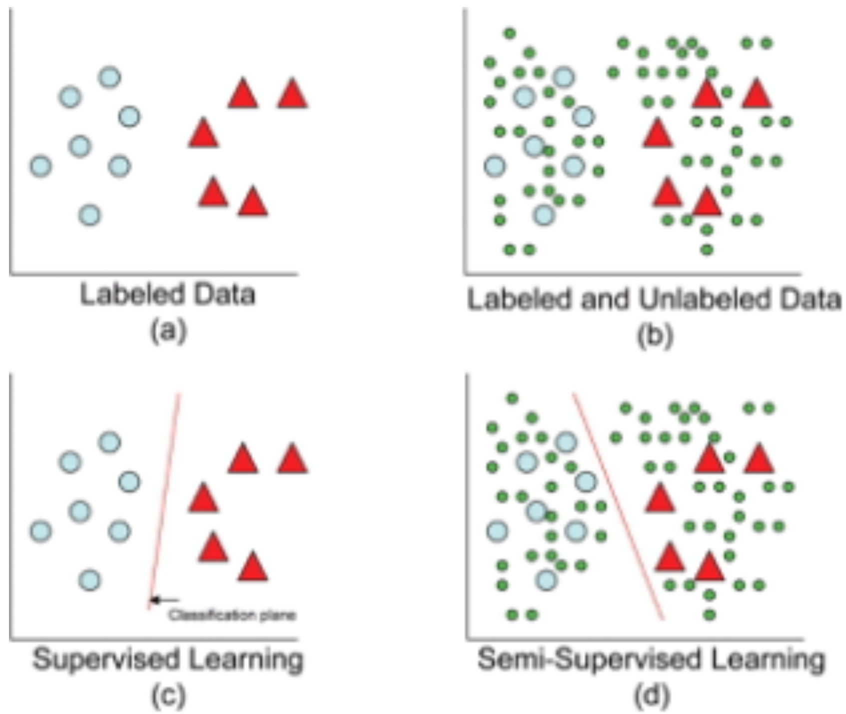
Algorithms	Applications
Kohonen's SOM	Phonetic typewriter, image compression, image segmentation, information management
K-means	Documents clustering, houses values identification, earthquake analysis
DBSCAN	Text clustering, climate analysis, earthquake analysis
PCA	Data compression, risk management of interest rate, neuroscience
Auto Encoders	Speech denoising, image denoising, edge detection

Semi-supervised Learning

Combination of both supervised and unsupervised learning

Why do we combine them?

- Labeled data are expensive.
- Unlabeled data are relatively easier to obtain

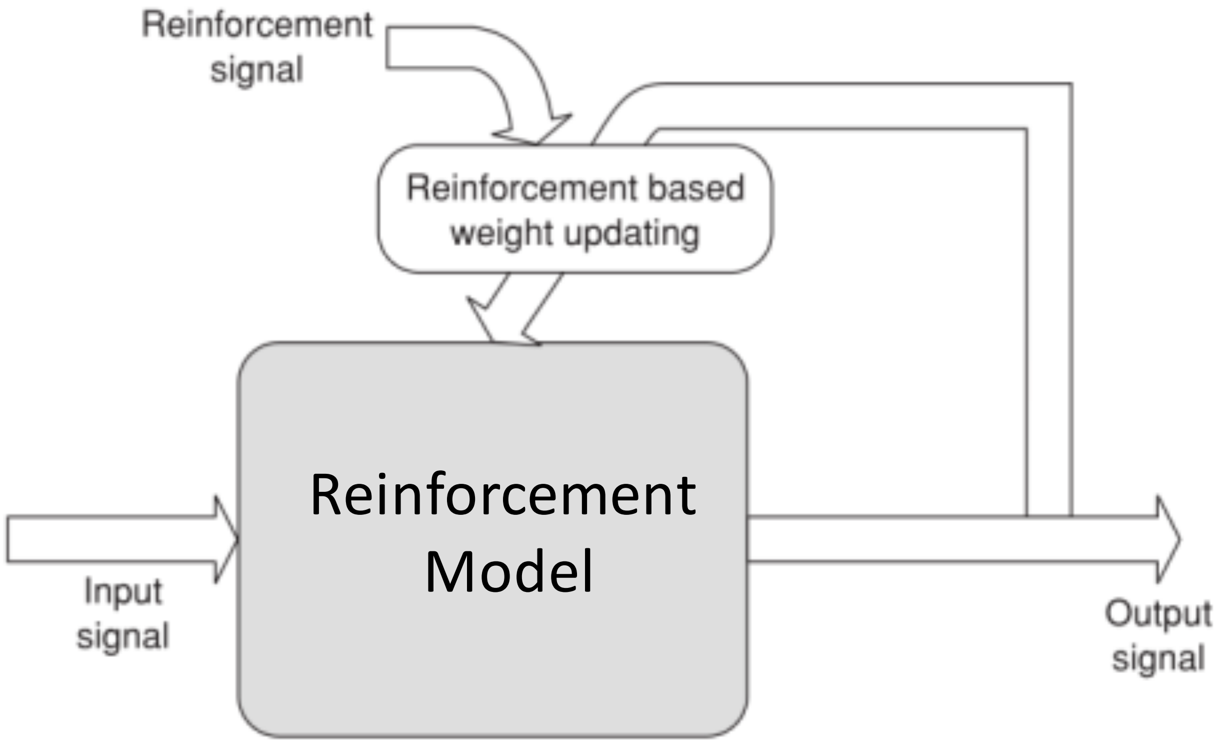


Well known method: Radial Basis Function (RBF) networks.

It is know the major approach of deep learning, e.g., Stacked RBMs to learn the features and then Back-propagation to fine-tune the

Reinforcement Learning

- Allows the machine or software agents to learn its behaviour based on feedback from the environment.
- This behaviour can be learnt once and for all, or keep on adapting as time goes by.



Generic Learning Process

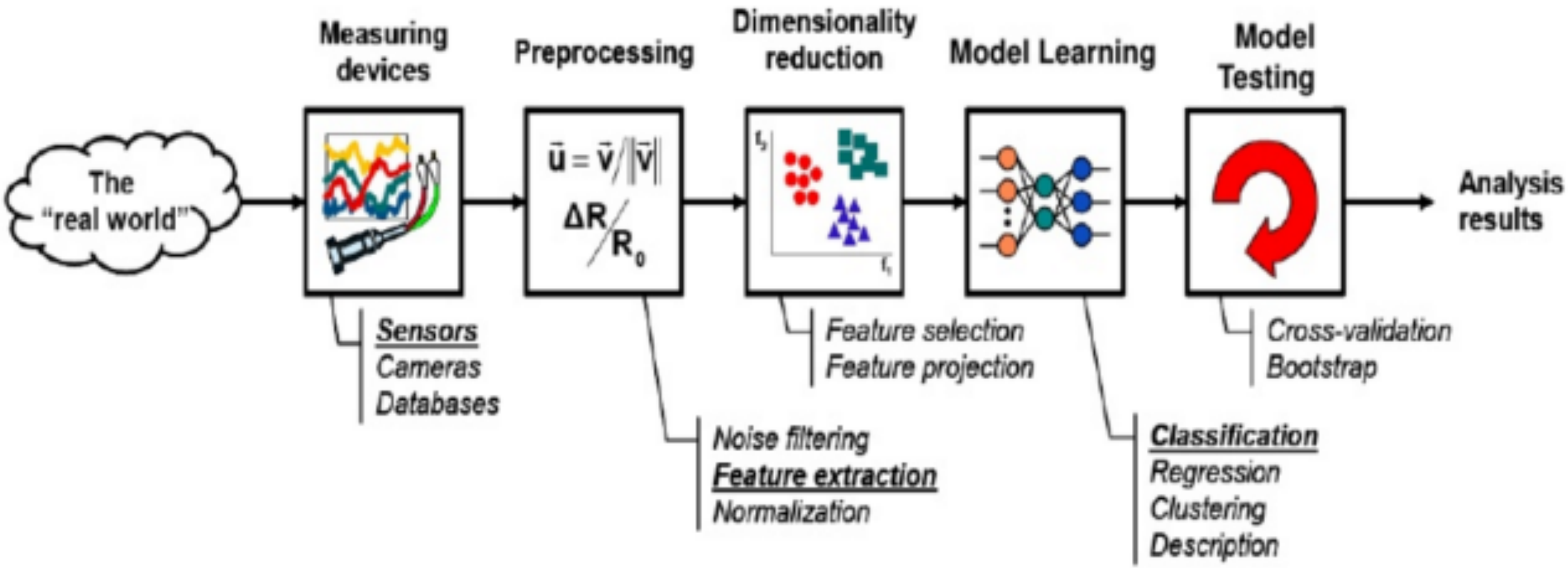


Image by Lior Rokach

Why Researchers Keep Developing New algorithms?

- There is no “best” algorithm for all applications. Some algorithms are appropriate only to certain applications.
- The types of application or problems are increasingly challenging.

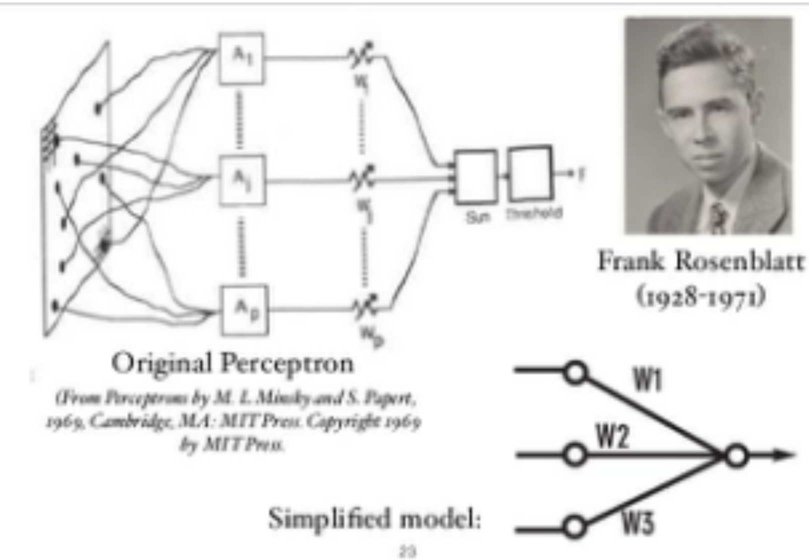
- Advances in Big Data storage and processing power of machines (Hadoop, MapReduce), allowed us to explore more complex machine learning tools such as **Deep Learning.**

History and State of the Art of Connectionist ML

- Early days of AI. Invention of artificial neuron [McCulloch and Pitts, 1943] & perceptron [Rosenblatt, 1958]
- AI Winter. [Minsky and Papert, 1969] showed perceptron only learns linearly separable concepts
- Revival in 1980s: Multi-layer Perceptrons (MLP) and Back-propagation [Rumelhart et al., 1986]
- Other directions (1990s - present): SVMs, Bayesian Networks
- Revival in 2006: Deep learning [Hinton et al., 2006]
- Successes in applications: Speech at IBM/Toronto [Sainath et al., 2011], Microsoft [Dahl et al., 2012]. Vision at Google/Stanford [Le et al., 2012]

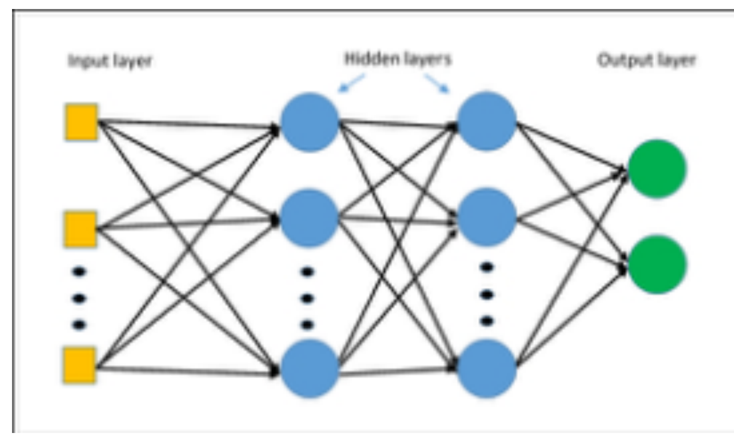
Perceptron and Multi-layer Perceptron

Perceptron



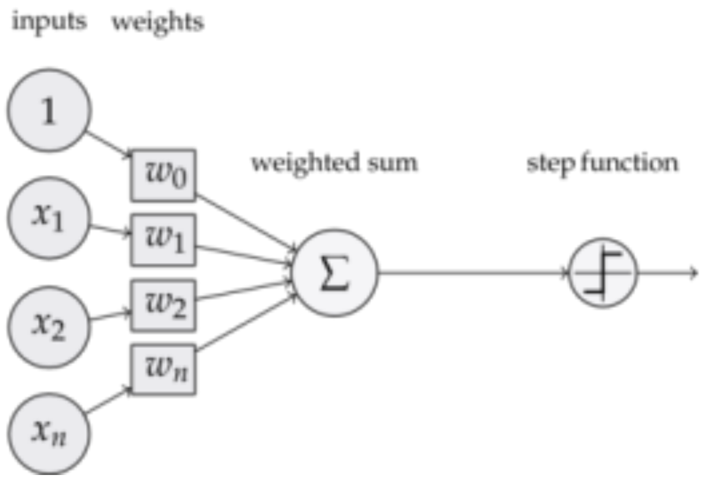
Multi-layer Perceptron

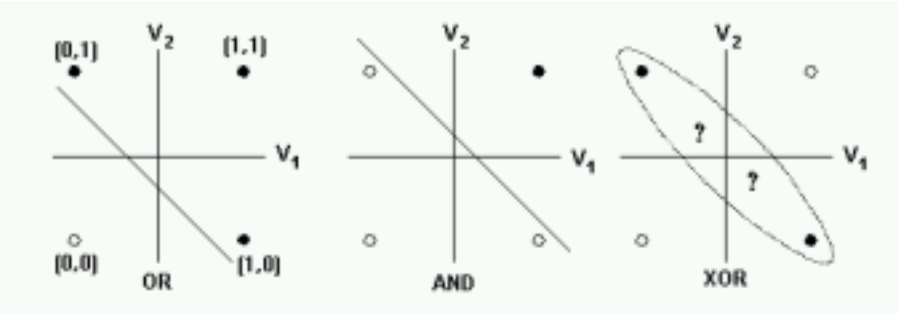
Rumelhart et. al., 1986



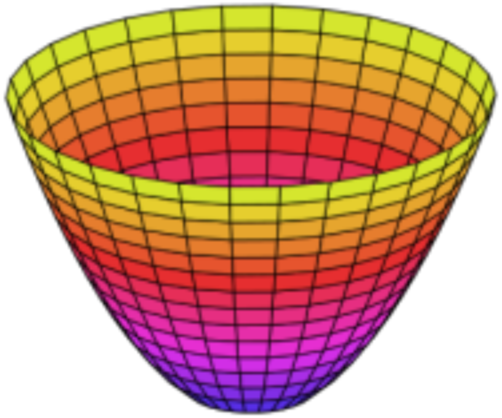
Nonlinear model

Perceptron



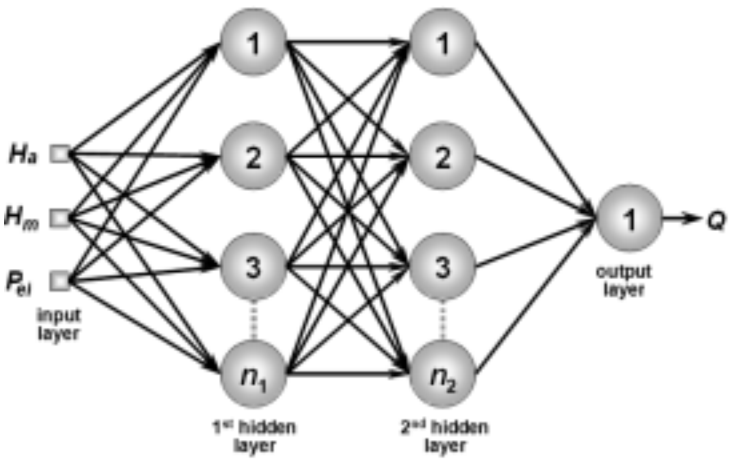


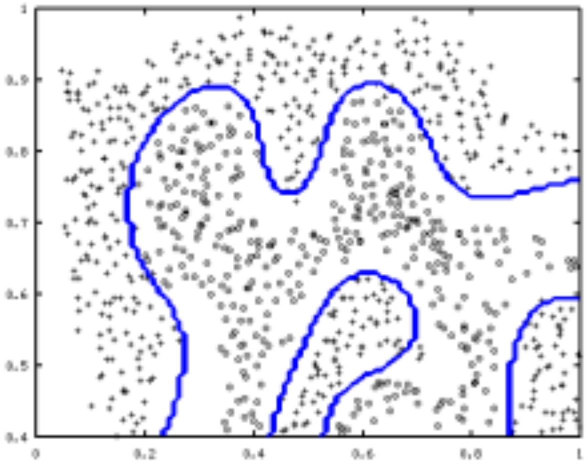
Error surface of a linear neuron



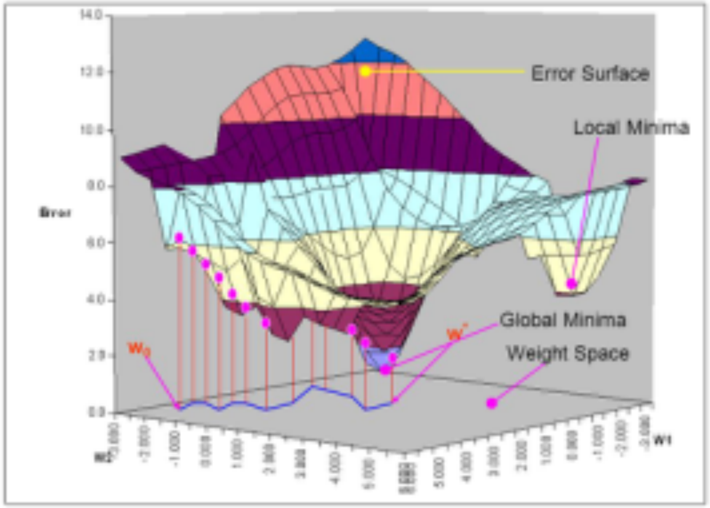
- Suitable for :
  - –linearly separable data
  - –very simple two-class Classification problem
- Does not generalize well.
- Uses a hard threshold function.

Multi-layer Perceptron





Error surface of an mlp



- Suitable for :
  - –non-linearly separable data
  - –complex Classification problem
- Uses a hard threshold function.
- Easy to overfit the data
- Optimization issues
- Difficult to train when several layers are used

Limitations of Shallow Networks

- The Curse of Dimensionality
- Insufficient model expressiveness
- Local Optimum Issue
- Backpropagation becomes ineffective due to vanishing gradients
- Solution: **Deep Architectures with Big Data!**