Introduction to Artificial Neural Networks

Nguyen Ngoc Thao

Department of Computer Science, FIT
University of Science, VNU-HCM

Introduction to Artificial Neural Networks

The victory of computer over human

- In 1997, IBM Deep Blue won the reigning world chess champion, Gary Kasparov.

- In 2016, AlphaGo beat the 9-dan Go professional Lee Sedol by using deep learning without handicaps.

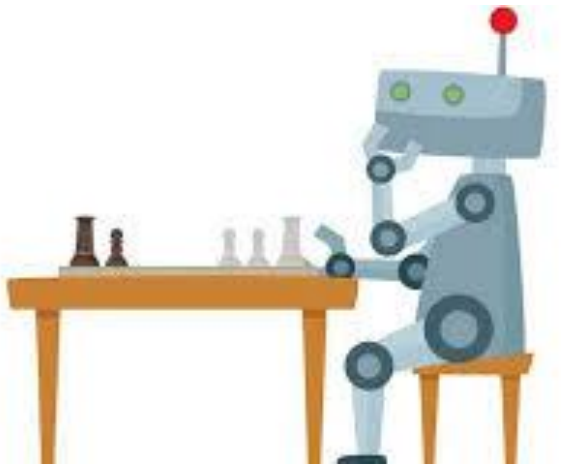
- In the future (2017), AlphaZero for Go, chess and shogi.

These winnings have marked turning points in the field of intelligent machines.

The complexity of Chess and Go

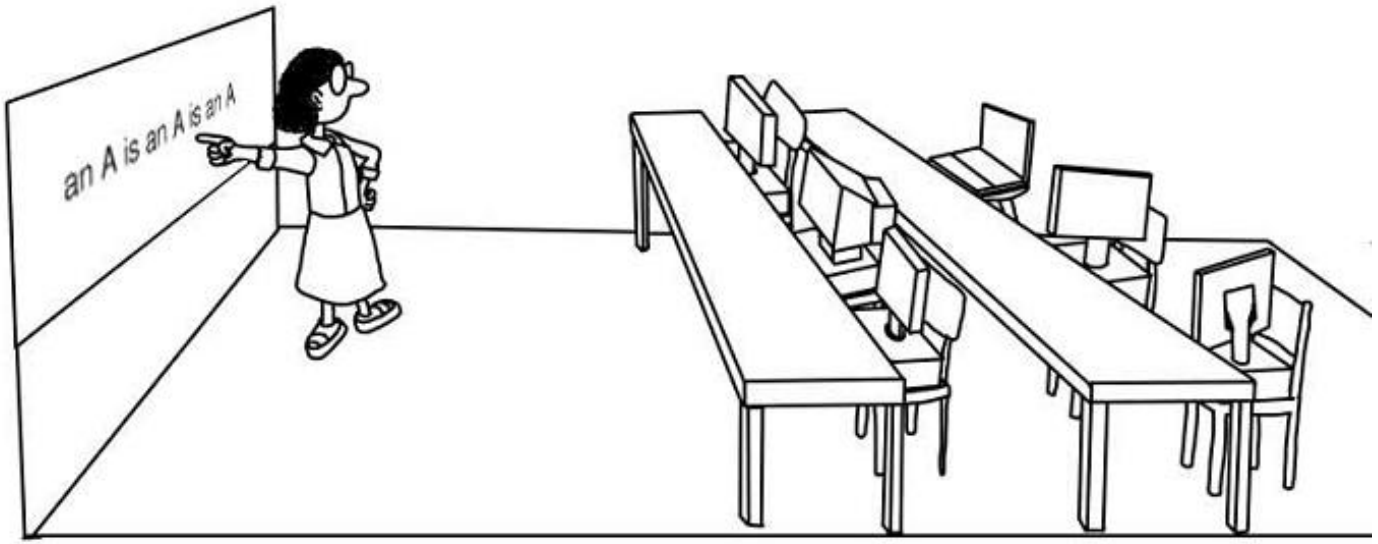
The incredible ability of computer

- Chess programs can formulate a strategy that goes beyond simply doing a great number of “look-ahead” moves/second.
  - E.g., Deep Blue – 200 million positions/sec vs. Kasparov – dozens of positions/second
- They must be able to improve performance with experience over time → learning capabilities

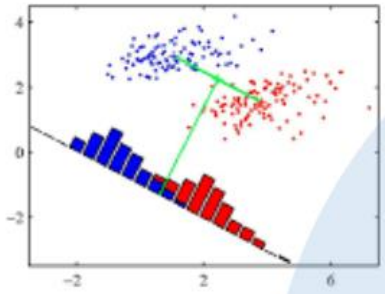


What is machine learning?

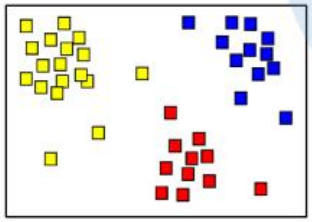
- Machine learning involves adaptive mechanisms that enable computers to learn from experience, learn by example and learn by analogy.

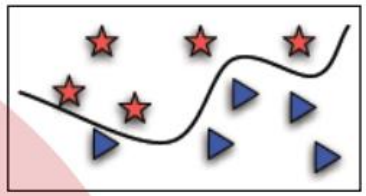


Types of machine learning

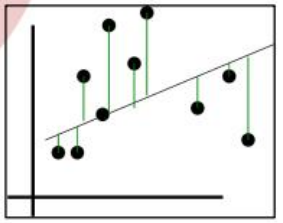


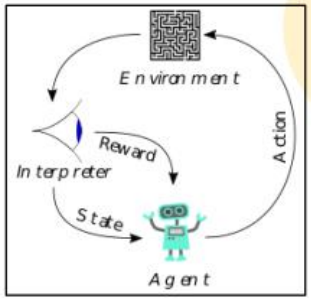
Unlabeled Data





Labeled Data





Multi-Step Problems

Machine Learning

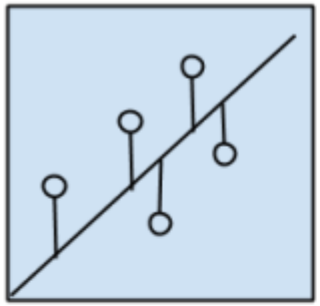
Unsupervised Learning
- Meaningful Compression
- Structure Discovery
- Dimensionality Reduction
- Big data Visualistaion
- Feature Elicitation
- Anomaly Detection
- Clustering
- Recommender Systems
- Targetted Marketing
- Association Mining
- Customer Segmentation

Supervised Learning
- Classification
- Image Classification
- Customer Retention
- Idenity Fraud Detection
- Diagnostics
- Feature Selection & Attribute Importance
- Regression
- Advertising Popularity Prediction
- Weather Forecasting
- Population Growth Prediction
- Market Forecasting
- Estimating life expectancy

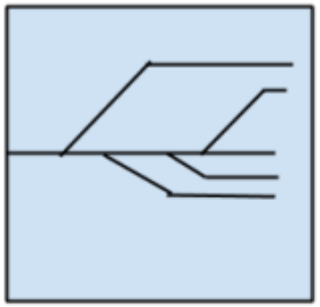
Reinforcement Learning
- Real-time decisions
- Game AI
- Robot Navigation
- Skill Acquisition
- Learning Tasks

Ref: https://ldi.upenn.edu/sites/default/files/Introduction-to-Machine-Learning.pdf

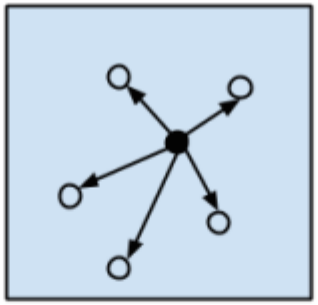
Machine learning algorithms



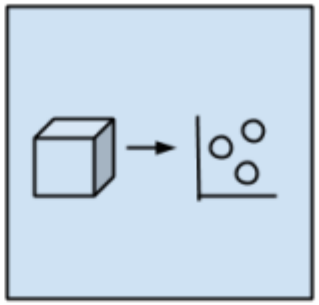
Regression Algorithms



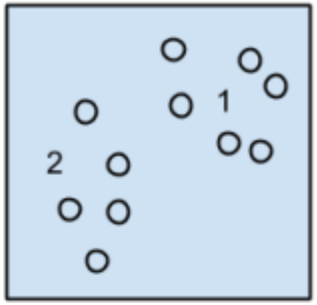
Regularization Algorithms



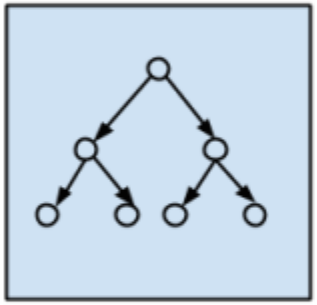
Instance-based Algorithms



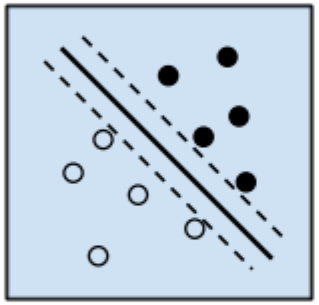
Dimensional Reduction Algorithms



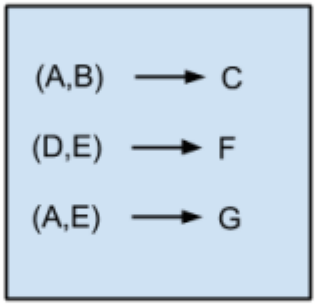
Clustering Algorithms



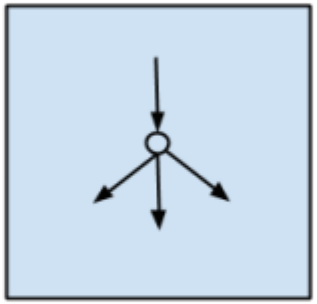
Decision Tree Algorithms



Support Vector Machines

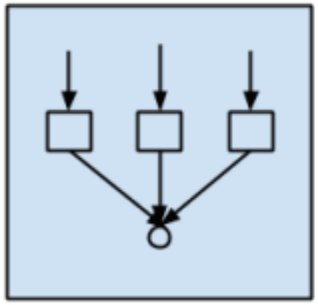


Association Rule Learning Algorithms



Artificial Neural Network Algorithms

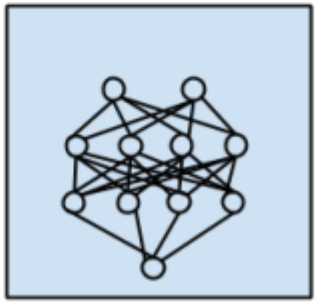
Bayesian Algorithms



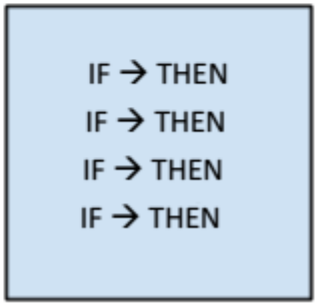
Ensemble Algorithms

Evolutionary Algorithms

Non-exhaustive list of ML families



Deep Learning Algorithms

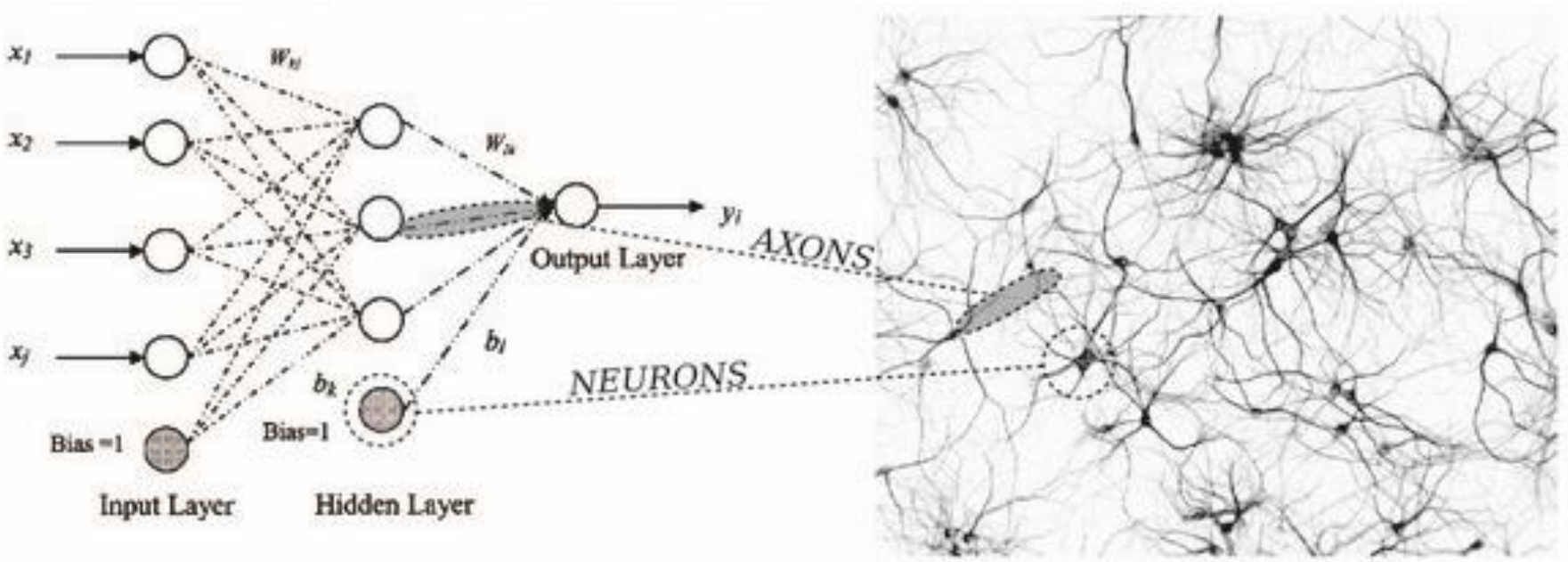


Learning Classifier Systems

Ref: https://ldi.upenn.edu/sites/default/files/Introduction-to-Machine-Learning.pdf

What is a neural network?

- A reasoning model based on the human brain, including billions of neurons and trillion connections between them

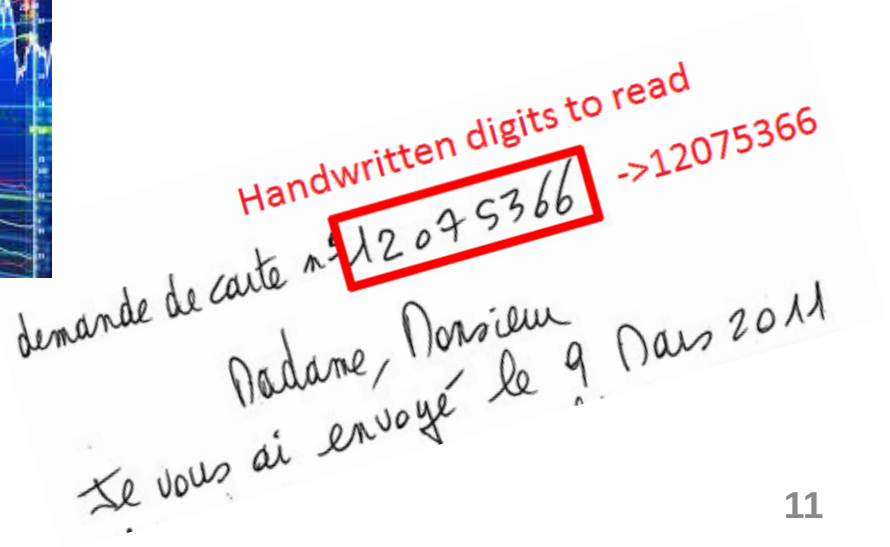


What is a neural network?

- The biological brain is considered as a highly complex, nonlinear and parallel information-processing system.
- Learning through experience is an essential characteristic.
- **Plasticity**: connections between neurons leading to the "right answer" are strengthened while those leading to the "wrong answer" are weaken.

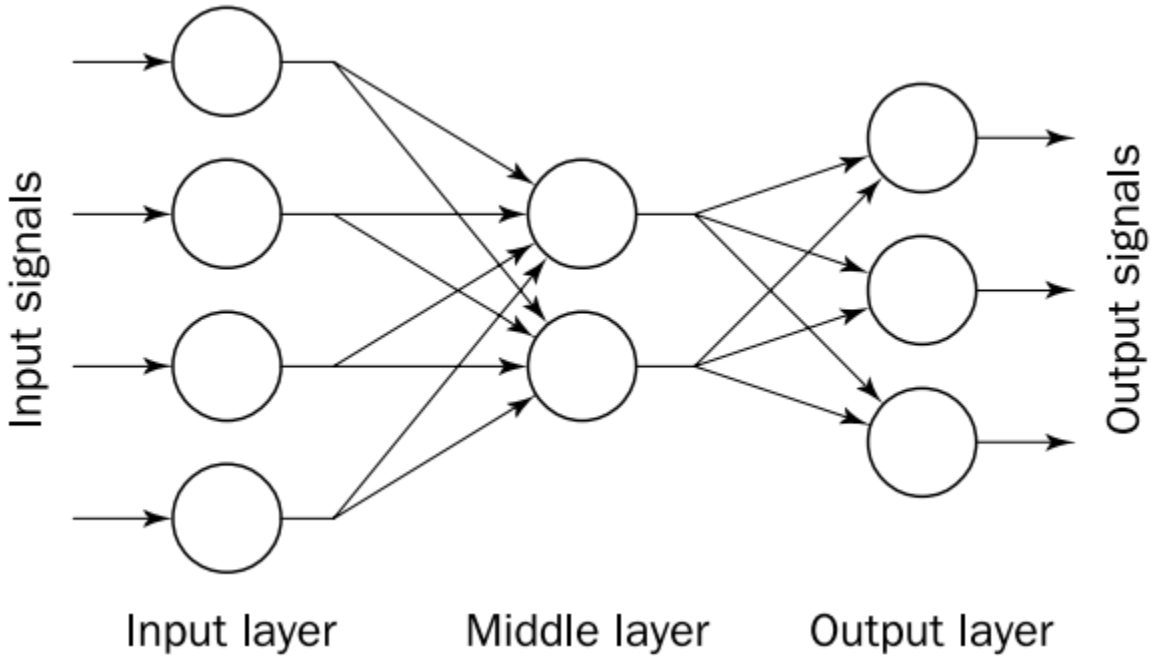
Artificial neural networks

- Artificial neural networks (ANNs) resembles the human brain in terms of learning mechanisms.
  - I.e., improve performance through experience and generalization
- Applications
  - Handwritten characters recognition, words identification in human speech, and explosives detection at airports, etc.



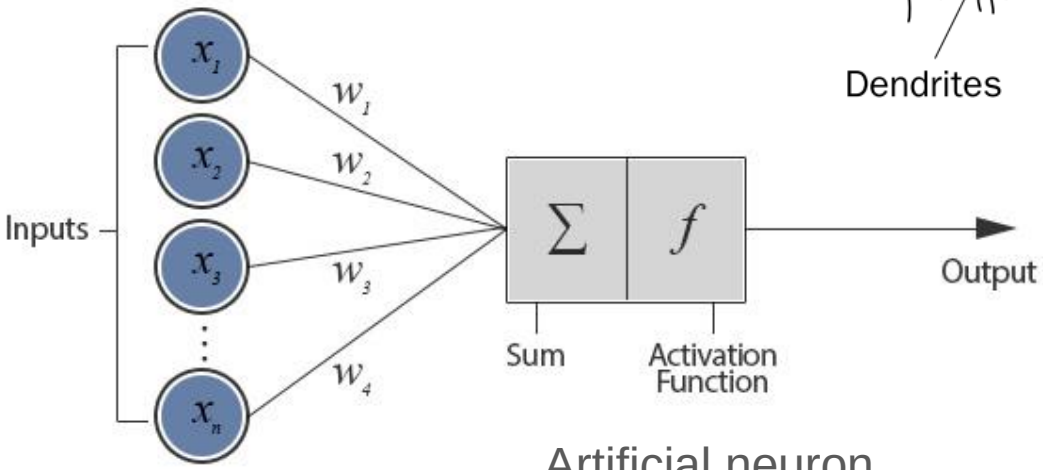
How do ANNs model the brain?

- An ANN consists of a number of very simple and highly interconnected processors (or neurons), arranging in a hierarchy of layers.

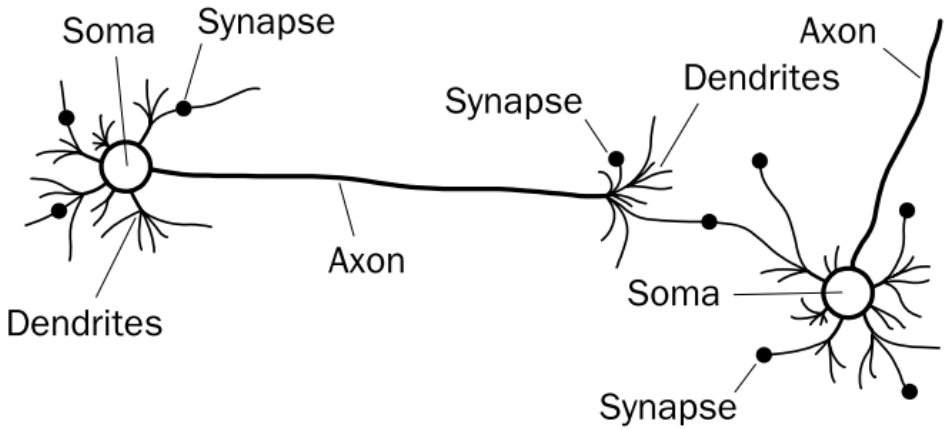


How do ANNs model the brain?

- Each neuron is an elementary information-processing unit.
- The neurons are connected by links passing signals from one neuron to another.
  - Each neuron receives a number of input signals through its connections and produces at most a single output signal.
- Each link associates with a numerical weight expressing the strength of the neuron input → basic means of long-term memory in ANNs
- ANNs “learn” through repeated adjustments of these weights



Artificial neuron



Biological neuron

Analogy between biological and artificial neural networks

Biological neural network	Artificial neural network
Soma	Neuron
Dendrite	Input
Axon	Output
Synapse	Weight

How to build an ANN?

- The network architecture must be decided first,
  - How many neurons are to be used? How the neurons are to be connected to form a network?
- Then determine which learning algorithm to use,
  - Supervised /semi-supervised / unsupervised / reinforcement learning
- And finally train the neural network
  - How to initialize the weights of the network? How to update them from a set of training examples.

Types of Artificial Neural Networks

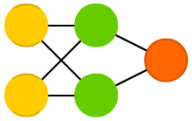
A mostly complete chart of

Neural Networks

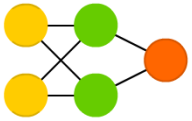
- Backfed Input Cell
- Input Cell
- Noisy Input Cell
- Hidden Cell
- Probablistic Hidden Cell
- Spiking Hidden Cell
- Output Cell
- Match Input Output Cell
- Recurrent Cell
- Memory Cell
- Different Memory Cell
- Kernel
- Convolution or Pool

Perceptron (P)

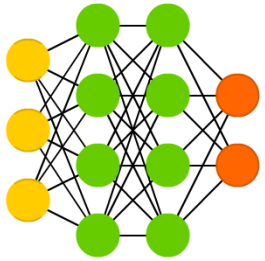
Feed Forward (FF)



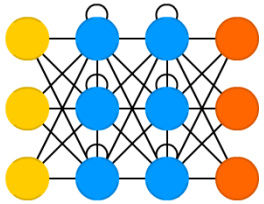
Radial Basis Network (RBF)



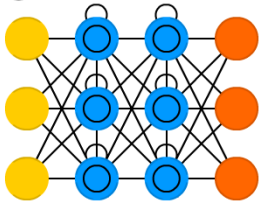
Deep Feed Forward (DFF)



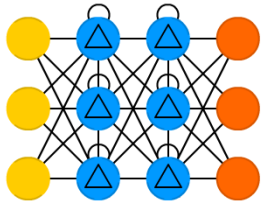
Recurrent Neural Network (RNN)



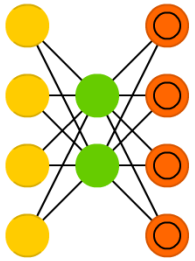
Long / Short Term Memory (LSTM)



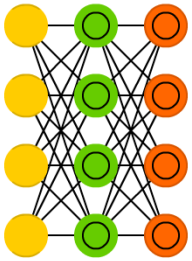
Gated Recurrent Unit (GRU)



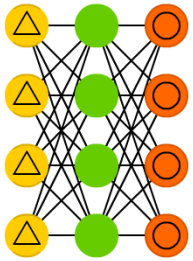
Auto Encoder (AE)



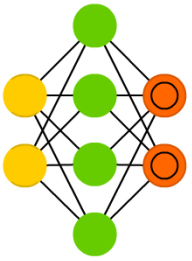
Variational AE (VAE)



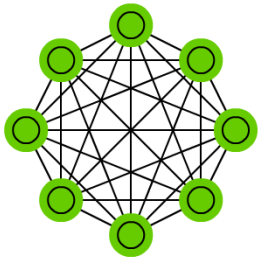
Denoising AE (DAE)



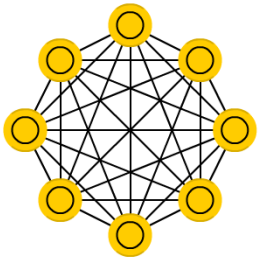
Sparse AE (SAE)



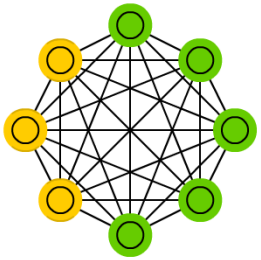
Markov Chain (MC)



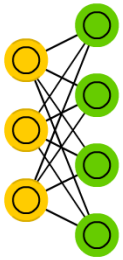
Hopfield Network (HN)



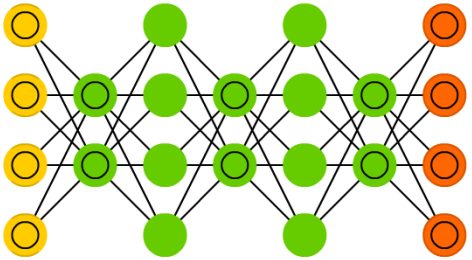
Boltzmann Machine (BM)



Restricted BM (RBM)



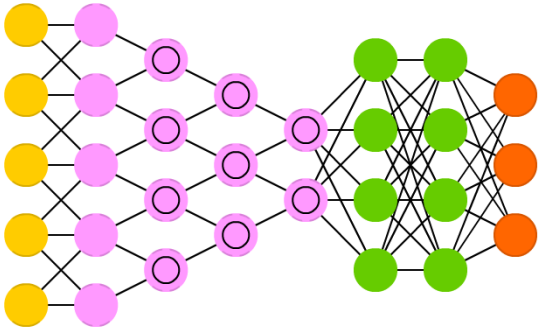
Deep Belief Network (DBN)



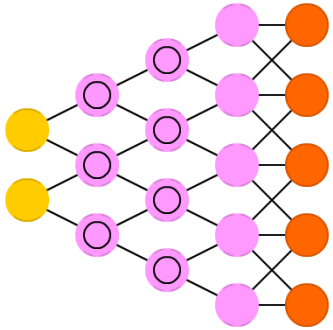
A mostly complete chart of

Neural Networks

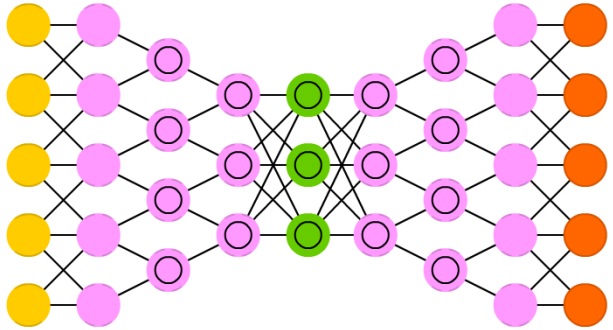
Deep Convolutional Network (DCN)



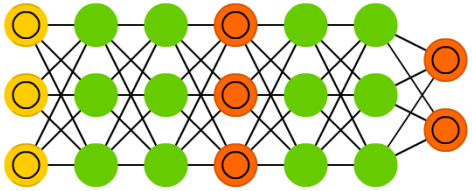
Deconvolutional Network (DN)



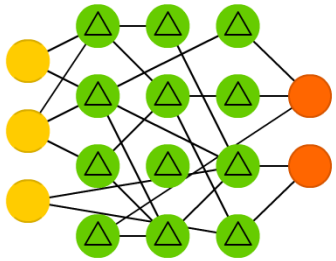
Deep Convolutional Inverse Graphics Network (DCIGN)



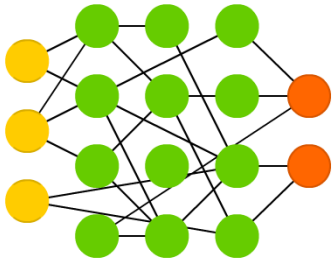
Generative Adversarial Network (GAN)



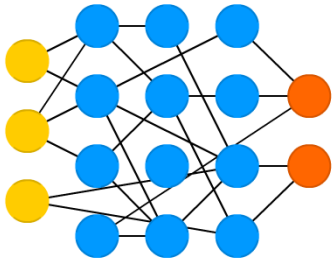
Liquid State Machine (LSM)



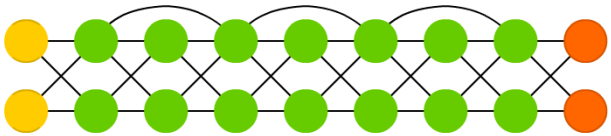
Extreme Learning Machine (ELM)



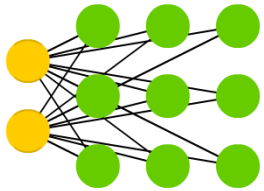
Echo State Network (ESN)



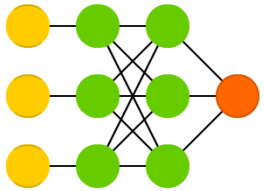
Deep Residual Network (DRN)



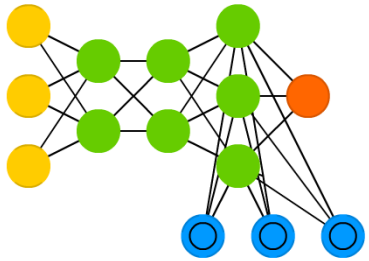
Kohonen Network (KN)



Support Vector Machine (SVM)



Neural Turing Machine (NTM)



Types of Learning Algorithms

Types of learning algorithms

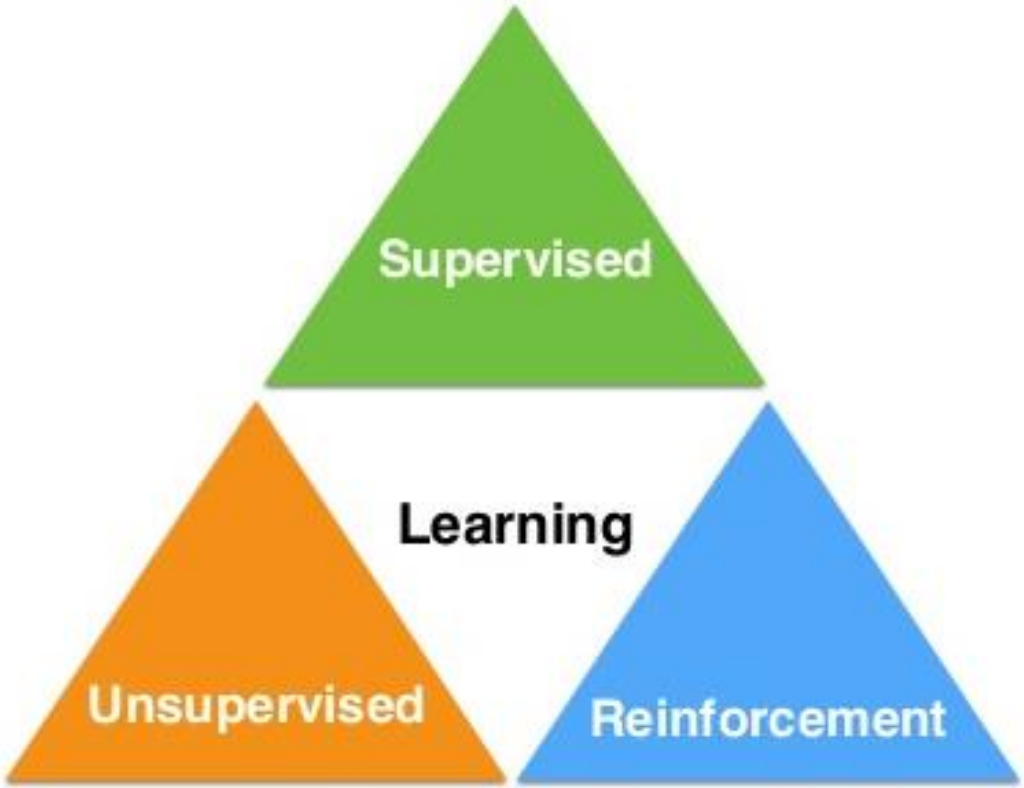
SUPERVISED LEARNING

UNSUPERVISED LEARNING

REINFORCEMENT LEARNING

Types of learning algorithms

- Labeled data
- Direct feedback
- Predict outcome/future

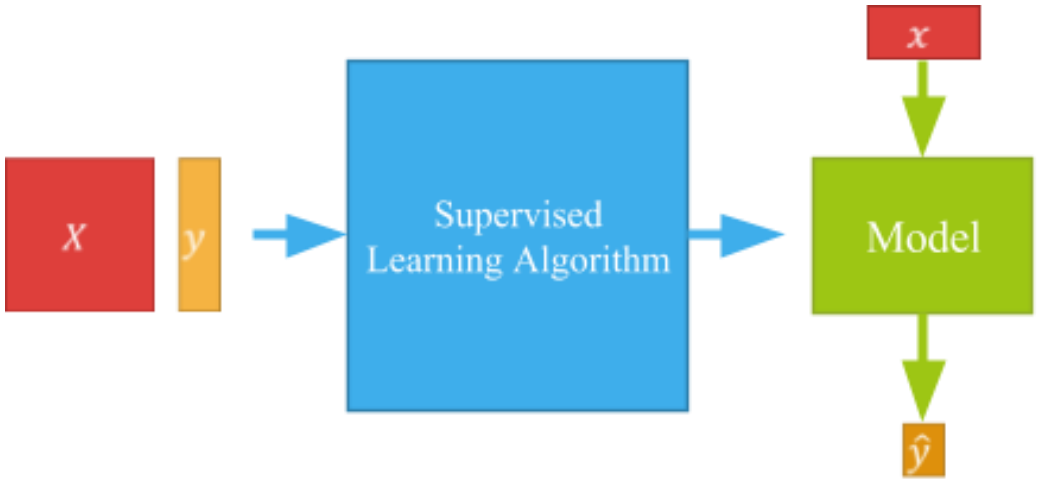


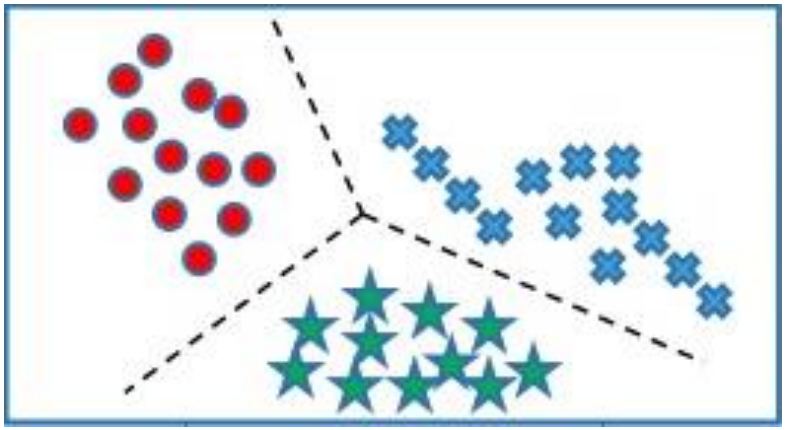
- No labels
- No feedback
- "Find hidden structure"

- Decision process
- Reward system
- Learn series of actions

Supervised learning

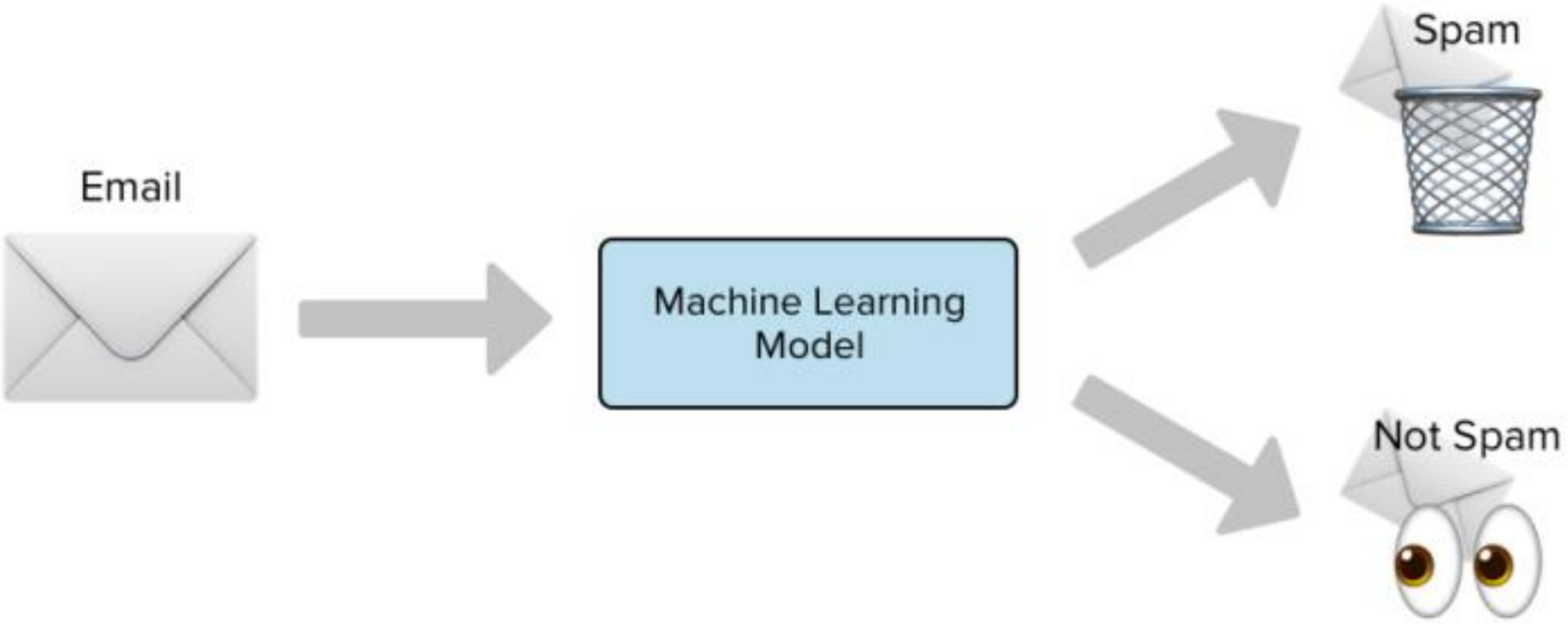
- Learn a function that maps an input to an output based on example input-output pairs





Supervised learning: Examples

- Spam detection: Decide which emails are spam and which are important
  - Use emails seen so far to obtain good prediction rule for future data



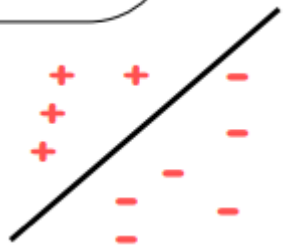
Supervised learning: Examples

- Spam detection: Decide which emails are spam and which are important
  - Represent each message by features. (e.g., keywords, spelling, etc.)

	"money"	"pills"	"Mr."	bad spelling	known-sender	spam?	
	Y	N	Y	Y	N	Y	
	N	N	N	Y	Y	N	
	N	Y	N	N	N	Y	
example	Y	N	N	N	Y	N	label
	N	N	Y	N	Y	N	
	Y	N	N	Y	N	Y	
	N	N	Y	N	N	N	

Reasonable RULES

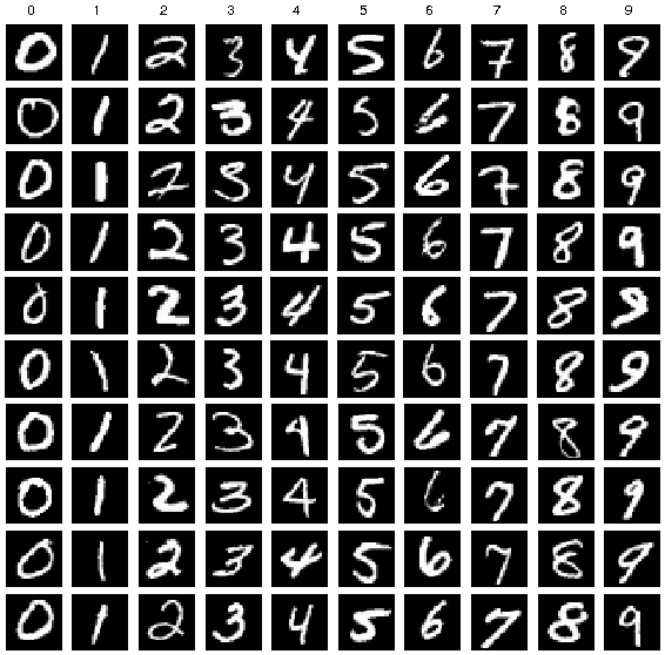
- Predict SPAM if unknown AND (money OR pills)
- Predict SPAM if 2money + 3pills – 5 known > 0



Linearly separable

Supervised learning: Examples

- Object detection and recognition: Localize and identify instances of semantic objects of a certain class (e.g., humans, buildings, or cars) in digital images and videos

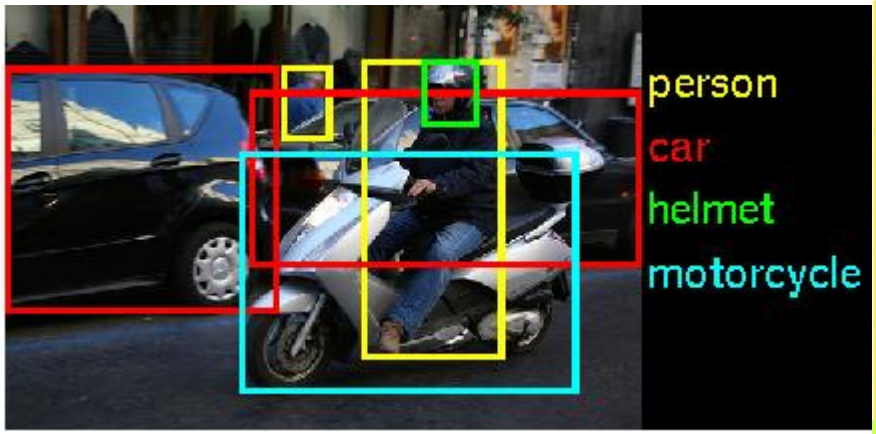


Handwritten digit recognition

Scene text recognition

Supervised learning: Examples

- Object detection and recognition: Localize and identify instances of semantic objects of a certain class (e.g., humans, buildings, or cars) in digital images and videos

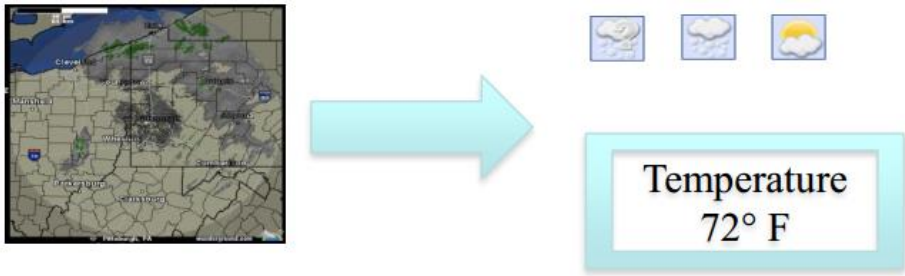


ImageNet object recognition

Indoor scene recognition

Supervised learning: More examples

- Weather prediction: Predict the weather type or the temperature at any given location…



- Medicine: diagnose a disease (or response to chemo drug X, or whether a patient is re-admitted soon?)
  - Input: from symptoms, lab measurements, test results, DNA tests, …
  - Output: one of set of possible diseases, or “none of the above”
  - E.g., audiology, thyroid cancer, diabetes, etc.

- Computational Economics:
  - Predict if a user will click on an ad so as to decide which ad to show
  - Predict if a stock will rise or fall (with specific amounts)

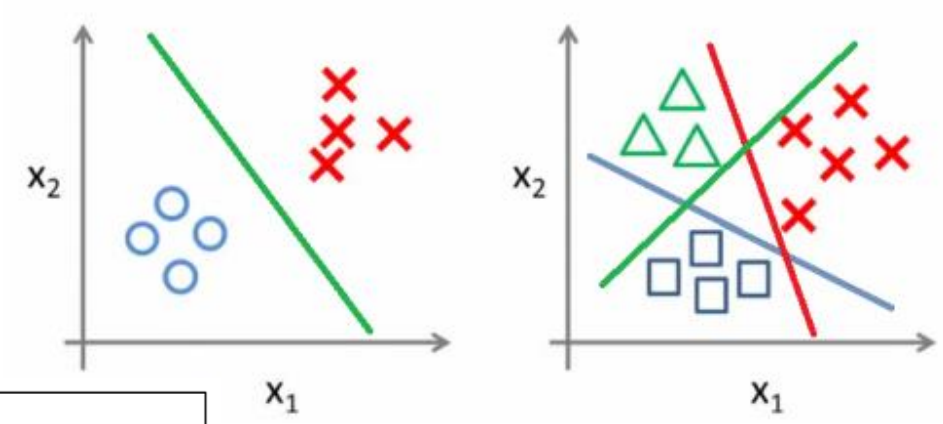
Classification vs. Regression

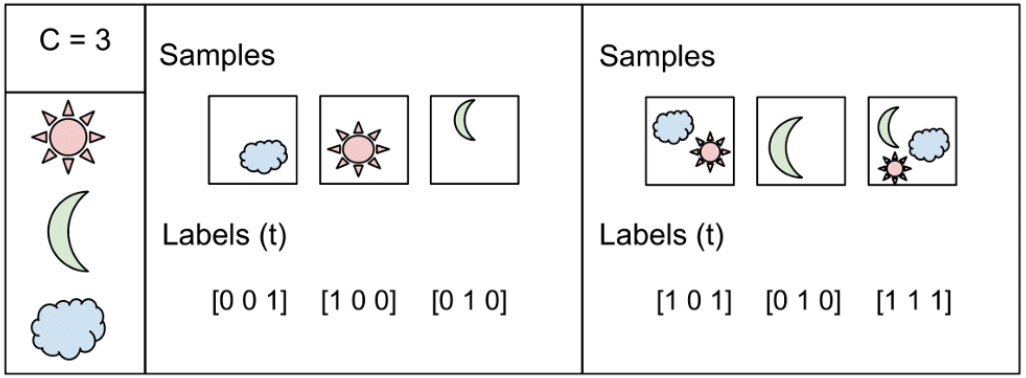
- Train a model to predict a categorical dependent variable
- Case studies: predicting disease, classifying images, predicting customer churn, buy or won't buy, etc.

- Binary classification vs.

  Multiclass classification vs.

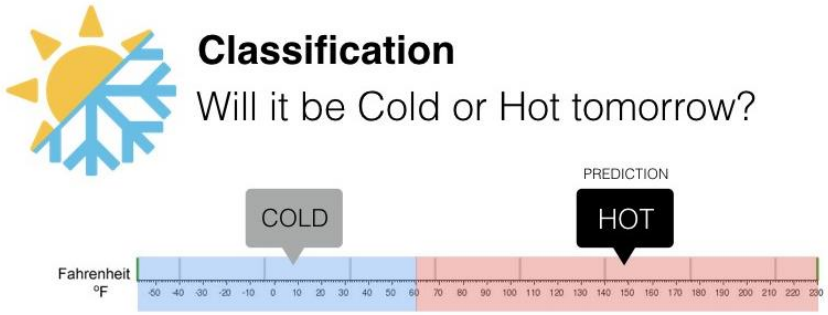
  Multilabel classification

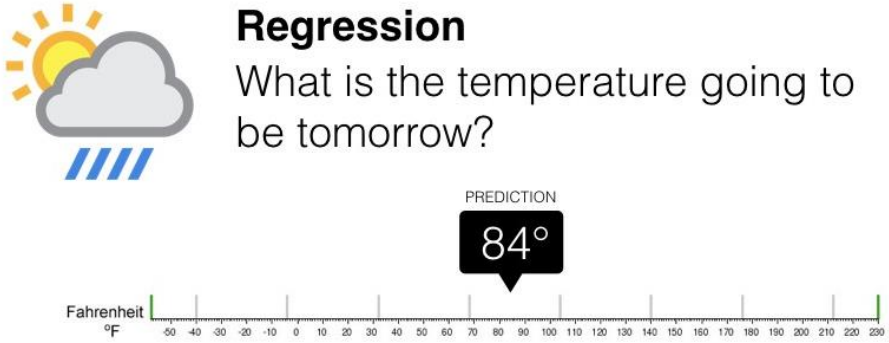




Classification vs. Regression

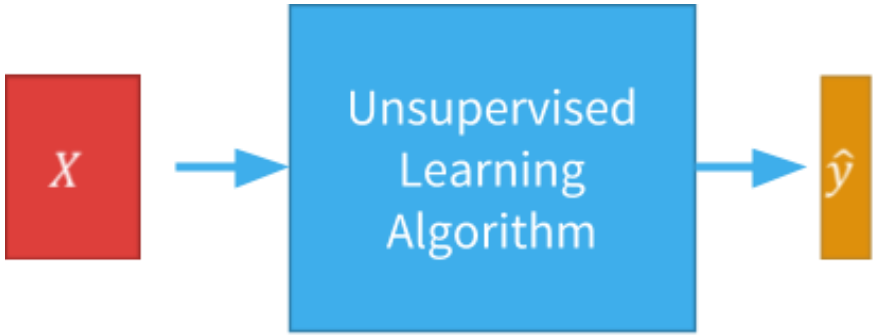
- Train a model to predict a continuous dependent variable
- Case studies: predicting height of children, predicting sales, forecasting stock prices, etc.

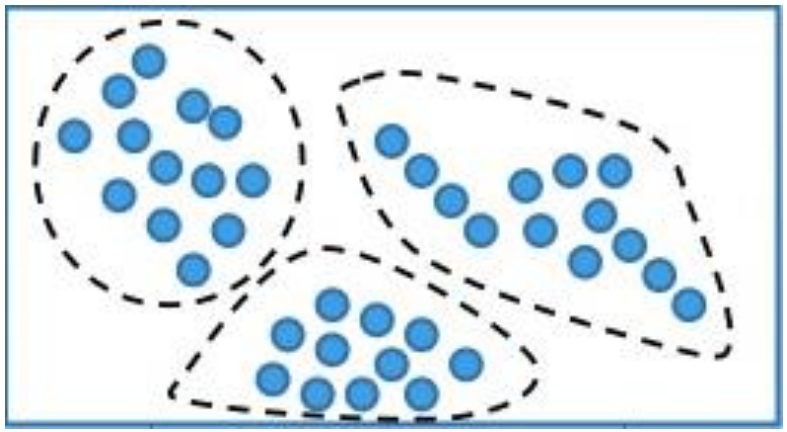




Unsupervised learning

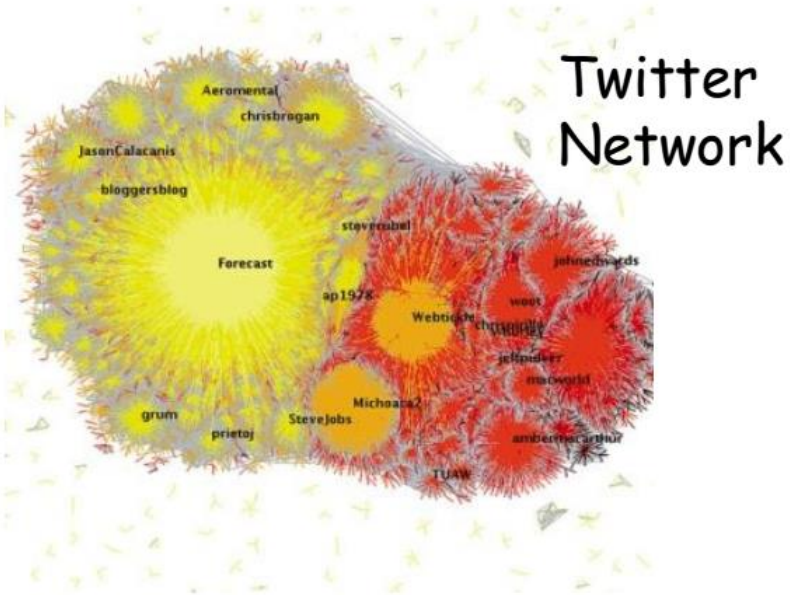
- Infer a function to describe hidden structure from "unlabeled" data
  - A classification (or categorization) is not included in the observations.

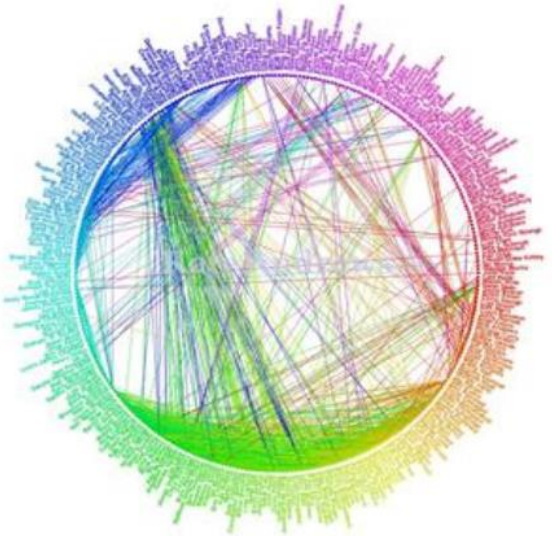




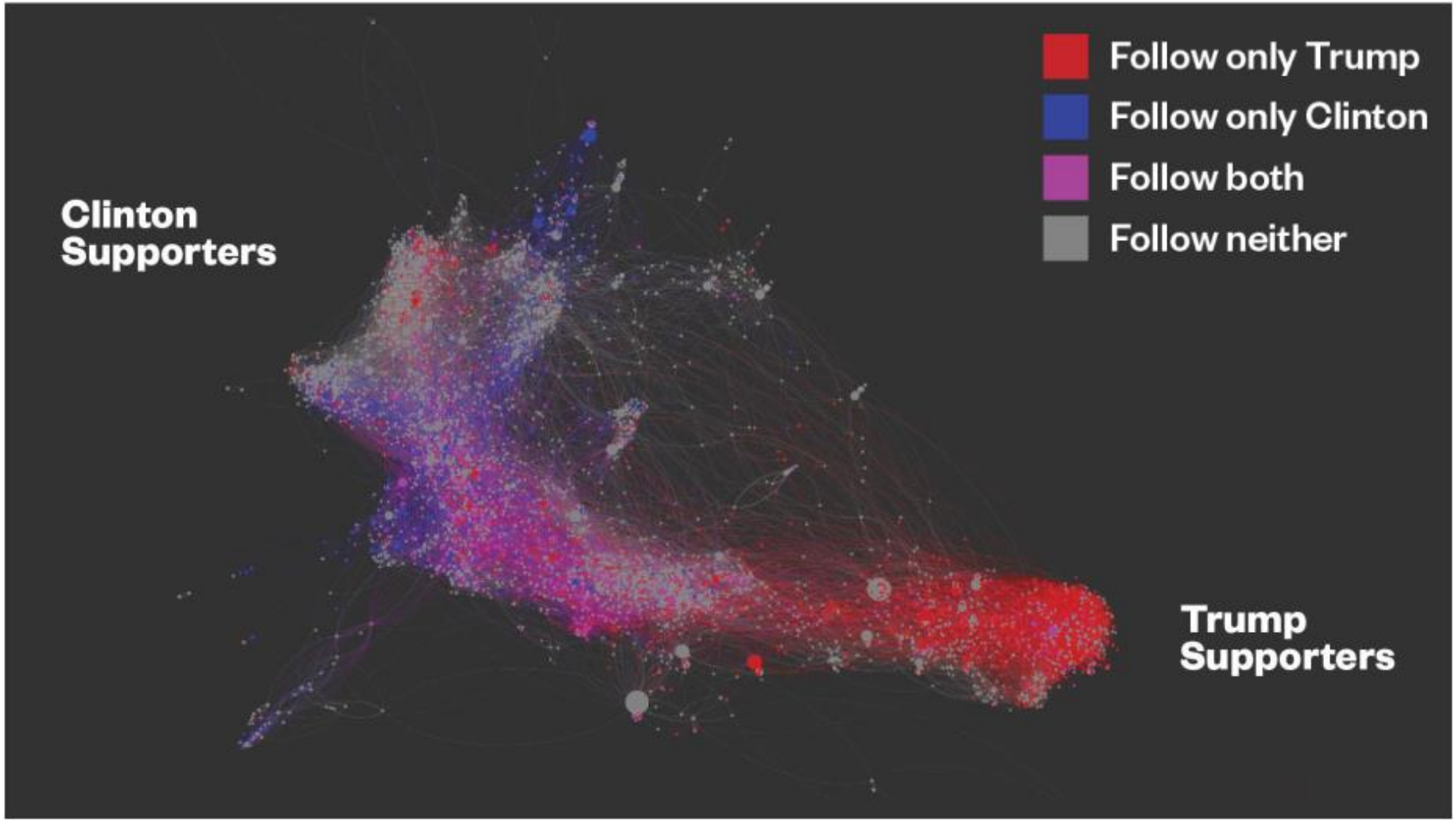
Unsupervised learning: Examples

- Social network analysis: cluster users of social networks by interest (community detection)



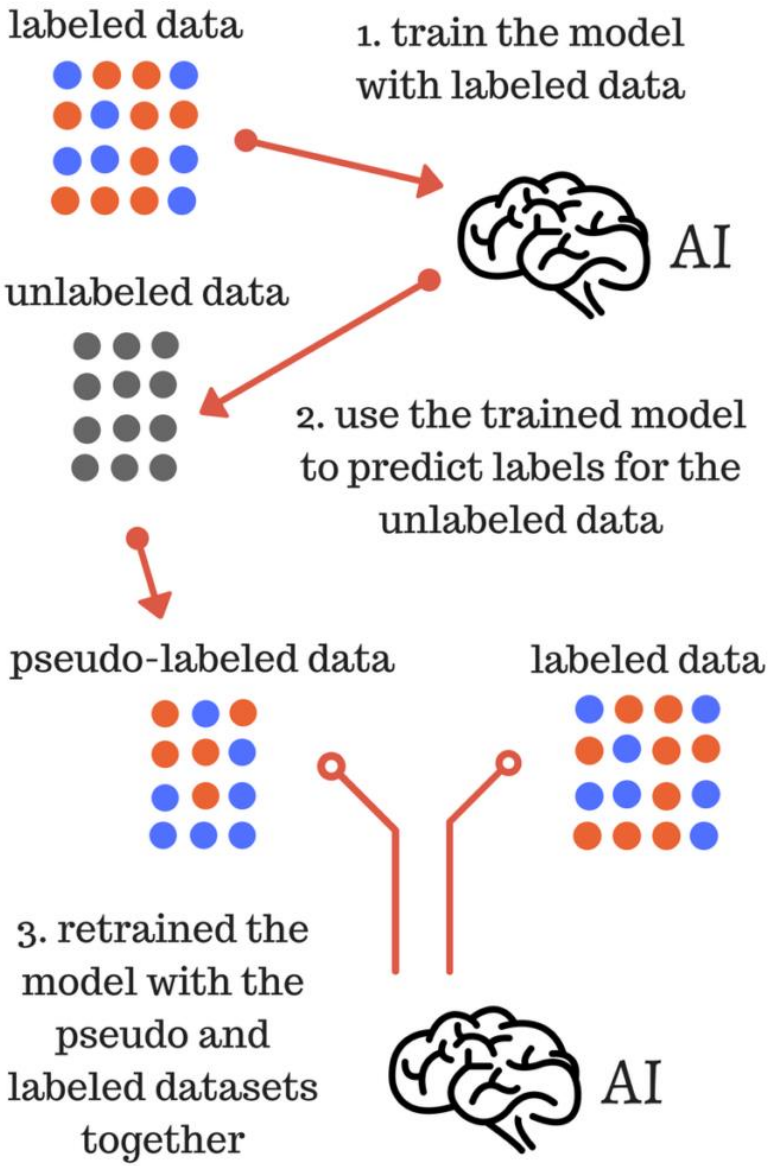


Unsupervised learning: Examples

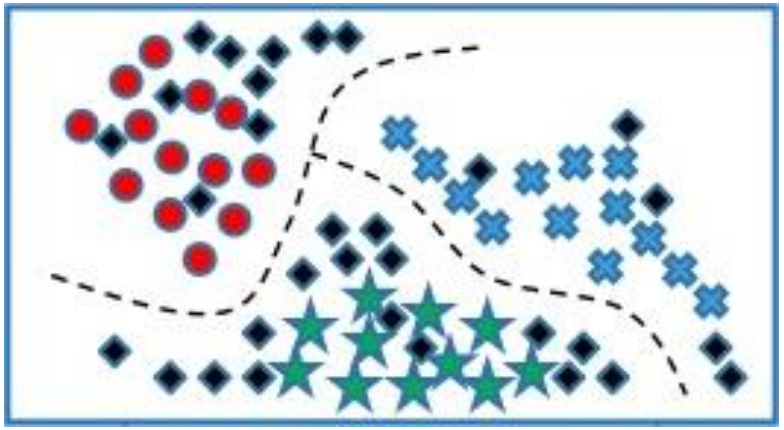


https://news.vice.com/en_us/article/d3xamx/journalists-and-trump-voters-live-in-separate-online-bubbles-mit-analysis-shows

Semi-supervised learning

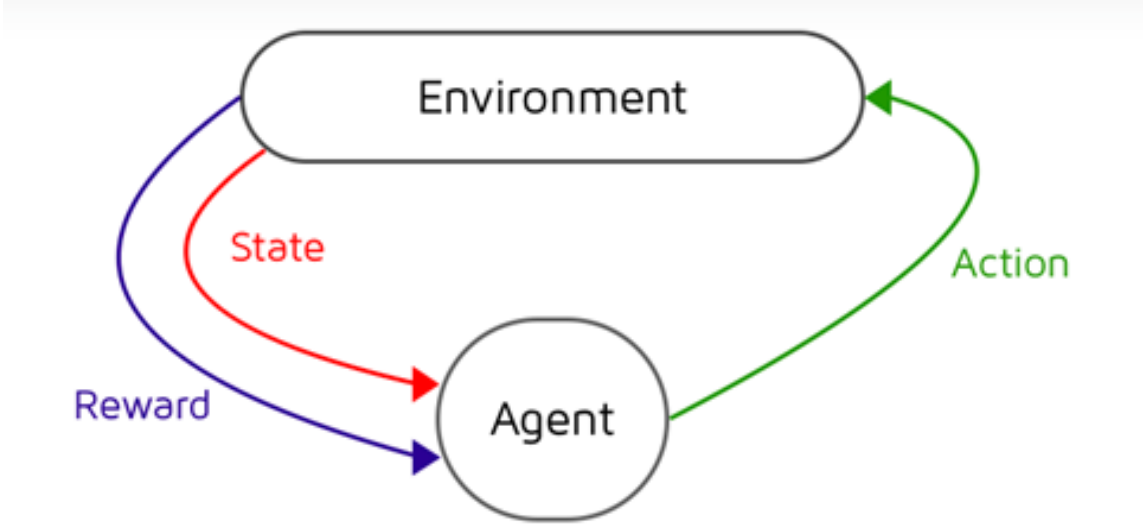


- The model is initially trained with a small amount of labeled data and a large amount of unlabeled data.

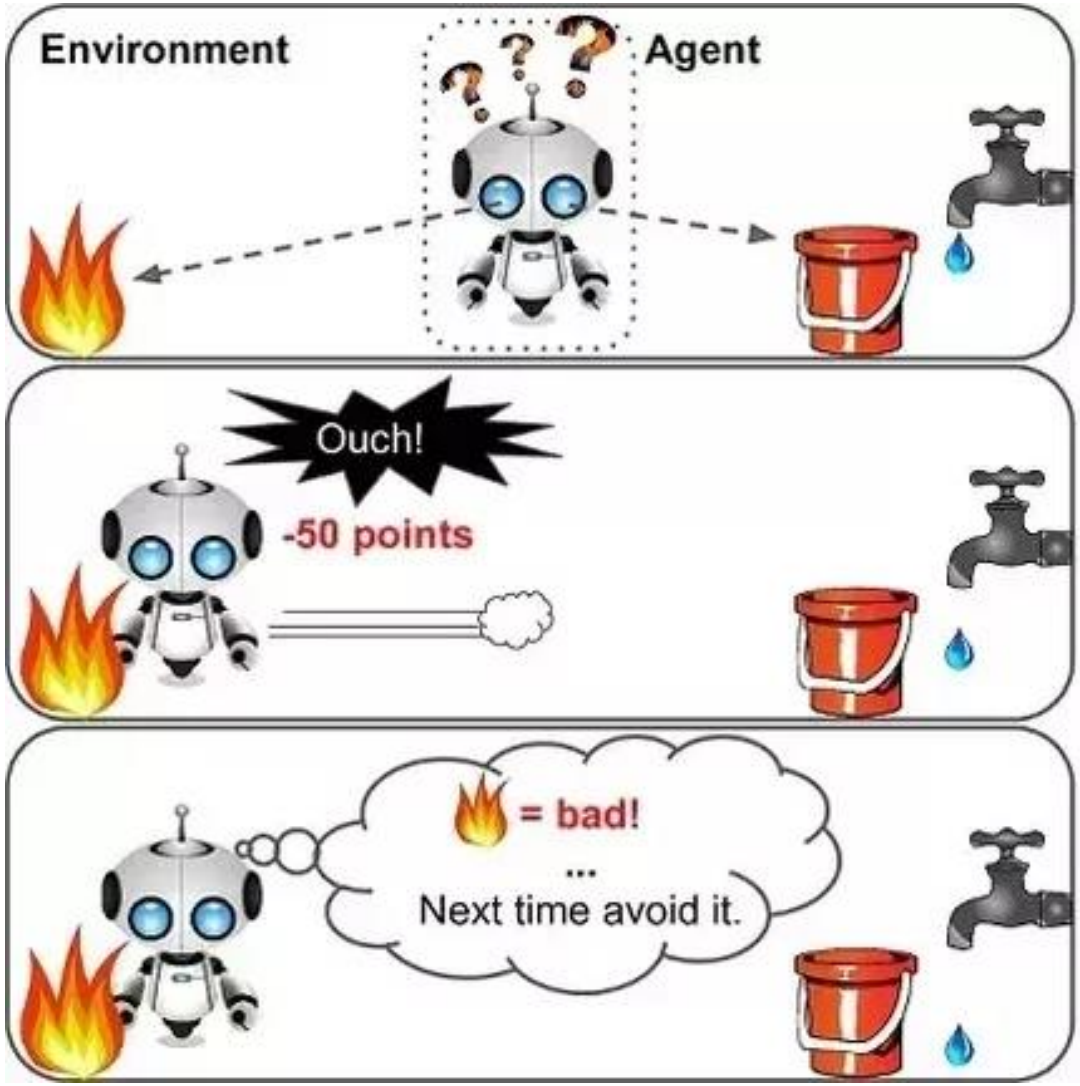


Reinforcement learning

- The agent learns from the environment by interacting with it and receives rewards for performing actions.

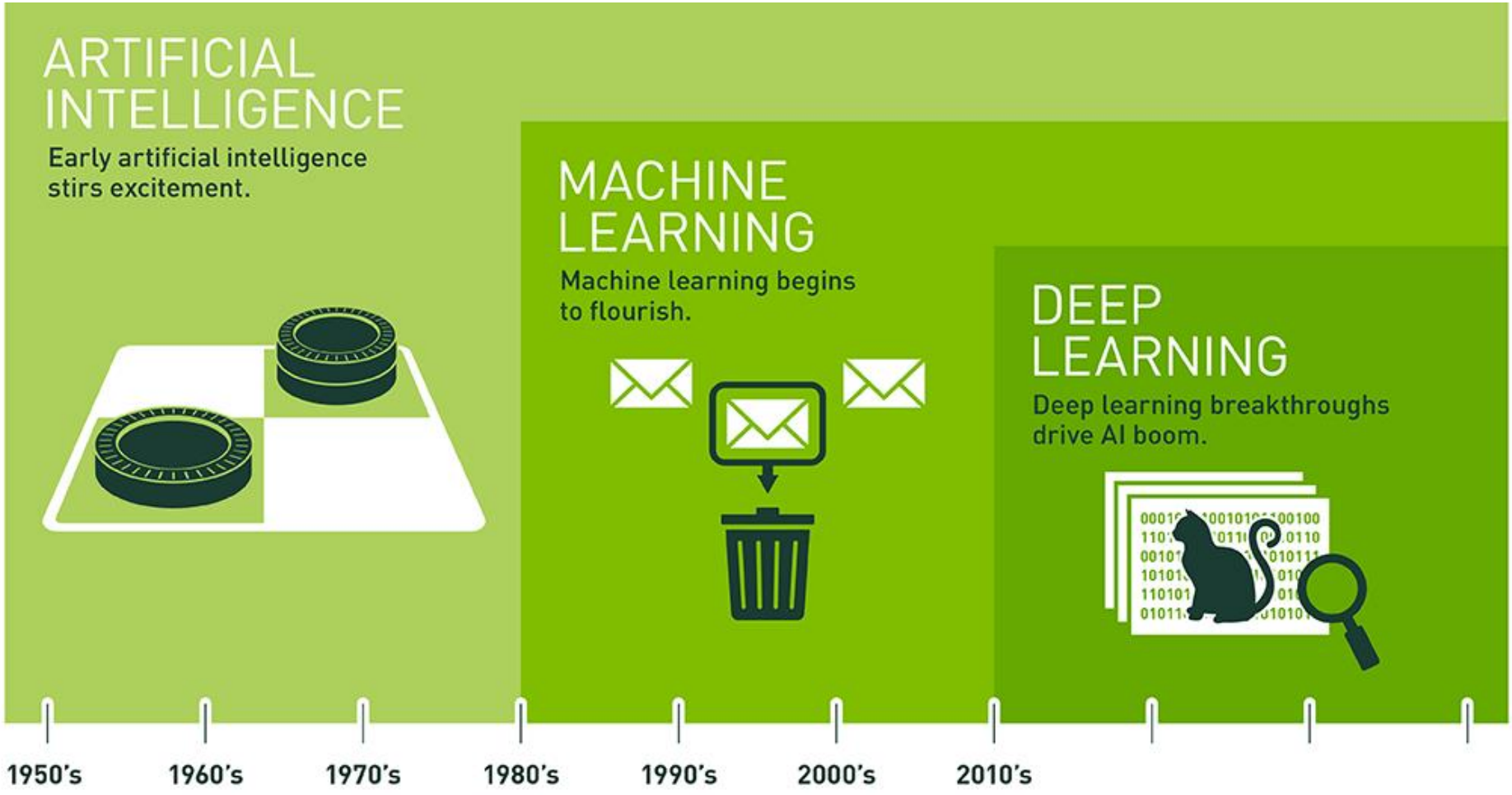


Reinforcement learning: Example



1. Observe
2. Select action using policy
3. Action!
4. Get reward or penalty
5. Update policy (learning step)
6. Iterate until an optimal policy is found

Machine learning and related concepts

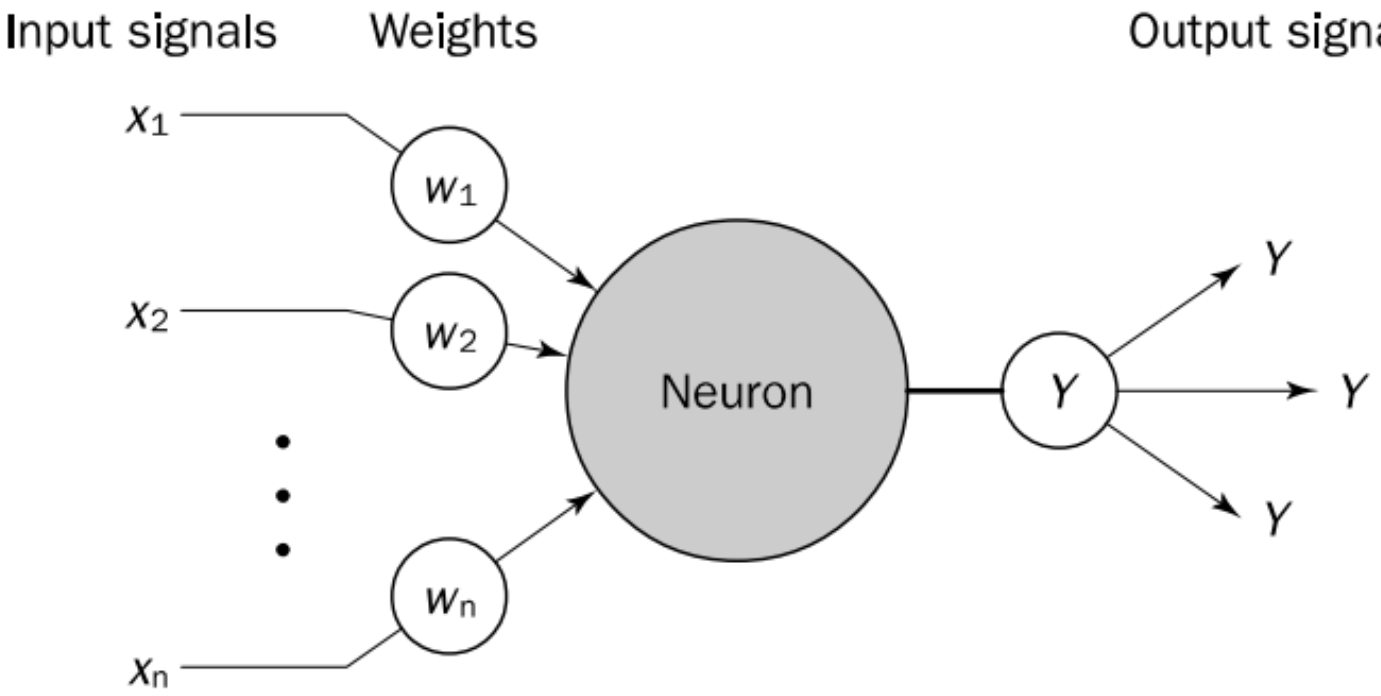


Source: https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

Neuron: An elementary information-processing unit in ANNs

Neuron: A simple computing element

- A neuron receives several signals from its input links, computes a new activation level and sends it as an output signal through the output links

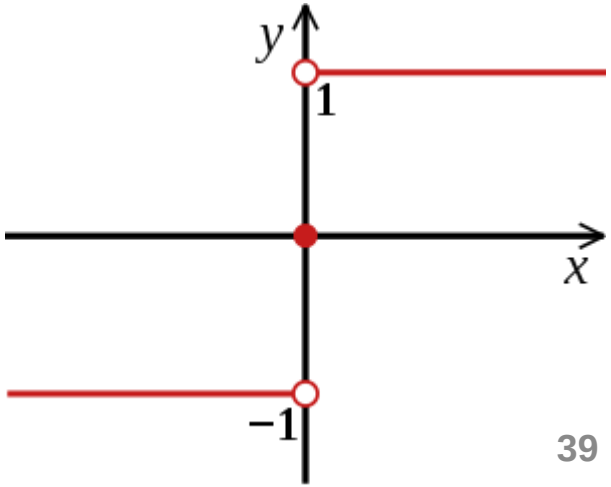


Neuron (McCulloch-Pitts, 1943)

- The weighted sum of input signals is passed to the activation function Y with a threshold θ

$$Y = \text{sign}\left[\sum_{i=1}^{n} x_i w_i - \theta\right]$$

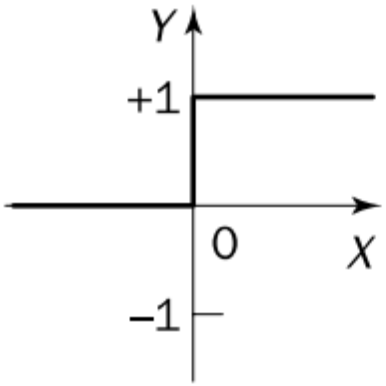
  - where x_i and w_i are the value of input i and its corresponding weight, respectively, n is the number of neuron inputs, and Y is the output value



Alternatives to activation function

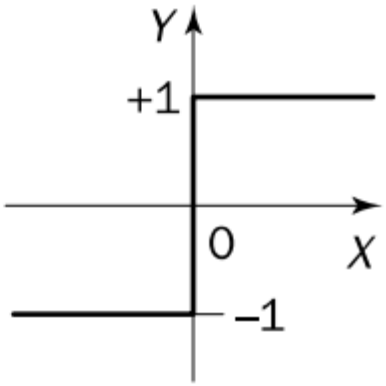
- Many activation functions have been tested, but only a few have found practical applications.

Step function



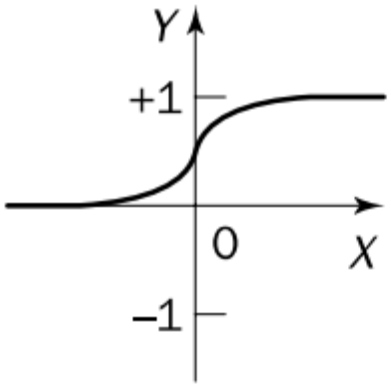
$$Y^{step} = \begin{cases} 1, \text{ if } X \geq 0 \\ 0, \text{ if } X < 0 \end{cases}$$

Sign function



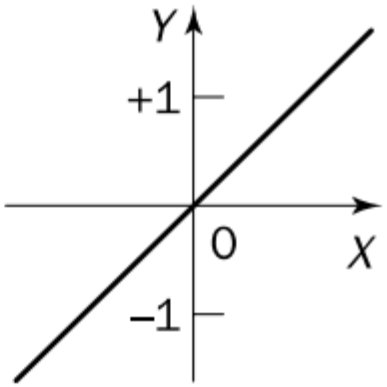
$$Y^{sign} = \begin{cases} +1, \text{ if } X \geq 0 \\ -1, \text{ if } X < 0 \end{cases}$$

Sigmoid function



$$Y^{sigmoid} = \frac{1}{1 + e^{-X}}$$

Linear function



$$Y^{linear} = X$$

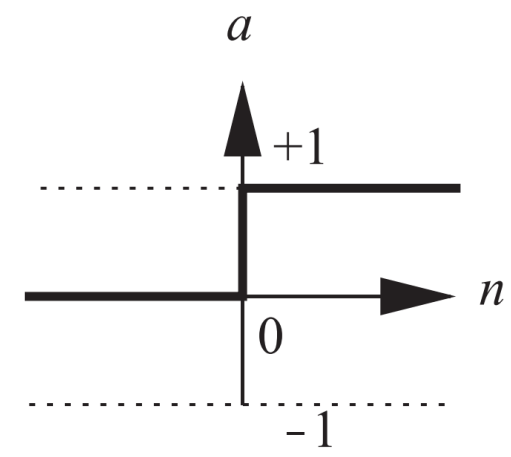
Types of Activation Functions

Activation functions

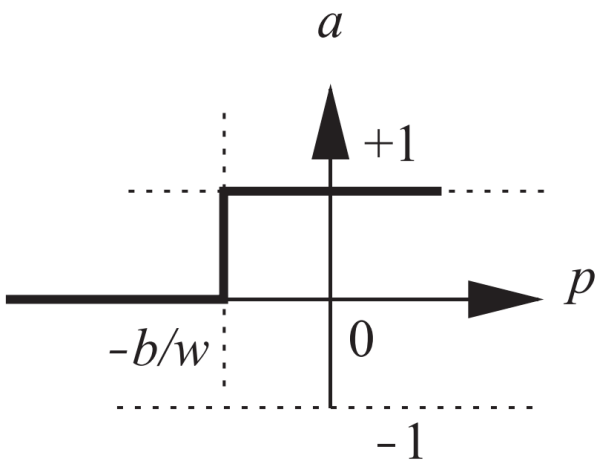
- The activation function (or transfer function) may be a linear or a nonlinear function of n.
- A particular function is chosen to satisfy some specification of the problem that the neuron is attempting to solve

Hard limit transfer function

- The output of the neuron is set to 0 if the function argument is less than 0, or 1 otherwise.



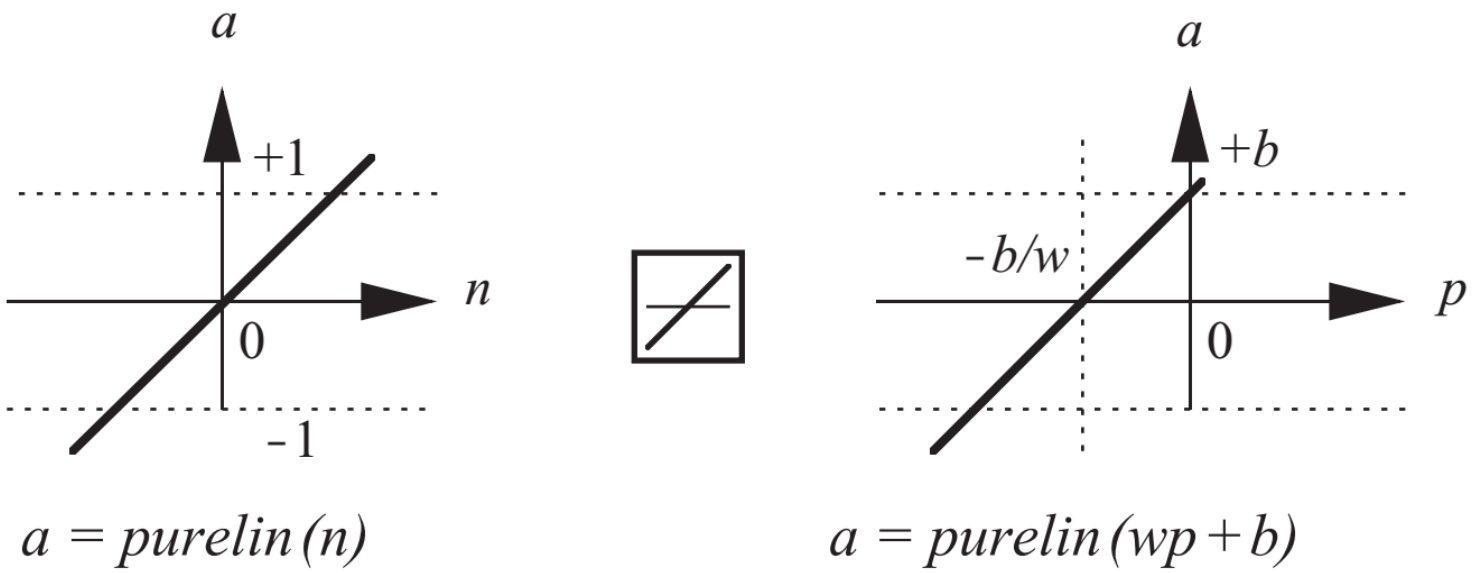
$a = hardlim(n)$



$a = hardlim(wp + b)$

Linear transfer function

- The output of the neuron is equal to its input, $a = n$.



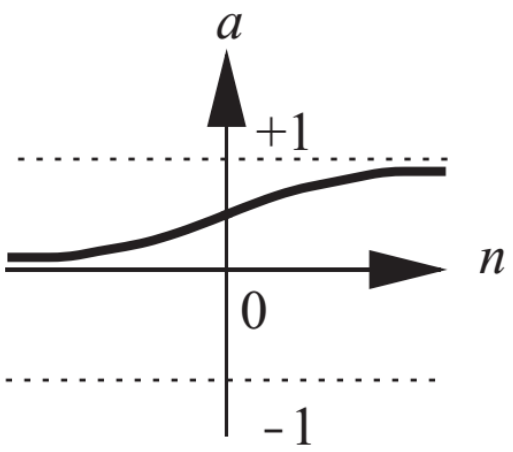
$a = purelin(n)$ $a = purelin(wp + b)$

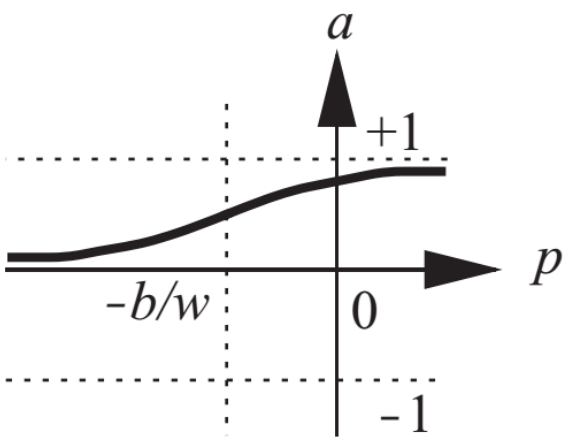
- Neurons with this transfer function are used in the ADALINE networks.

Log-sigmoid transfer function

- The function takes the input (which may have any value $\in [-\infty, +\infty]$) and squashes the output into the range [0, 1]

$$a = \frac{1}{1 + e^{-n}}$$





$a = logsig(n)$ $\qquad$ $a = logsig(wp + b)$

- Commonly used in multilayer networks that are trained using the backpropagation algorithm
  - This in part because this function is differentiable.

Transfer functions

Name	Input/Output Relation	Icon
Hard Limit	$a = 0 \quad n < 0$ $a = 1 \quad n \geq 0$	
Symmetrical Hard Limit	$a = -1 \quad n < 0$ $a = +1 \quad n \geq 0$	
Linear	$a = n$	
Saturating Linear	$a = 0 \quad n < 0$ $a = n \quad 0 \leq n \leq 1$ $a = 1 \quad n > 1$	
Symmetric Saturating Linear	$a = -1 \quad n < -1$ $a = n \quad -1 \leq n \leq 1$ $a = 1 \quad n > 1$	

Transfer functions

Name	Input/Output Relation	Icon
Log-Sigmoid	$a = \frac{1}{1+e^{-n}}$	
Hyperbolic Tangent Sigmoid	$a = \frac{e^{n}-e^{-n}}{e^{n}+e^{-n}}$	
Positive Linear	$a = 0 \quad n < 0$ $a = n \quad 0 \le n$	
Competitive	$a = 1$ neuron with max n $a = 0$ all other neurons	


THE
END