Neural Networks and Deep Learning

Ahmad Kalhor
Associate Professor .
School of Electrical and Computer Engineering –
University of Tehran
Spring 2023

Contents

1. Introduction

- Natural Neural Networks
- Artificial NNs and Applications
- Architectures, Activation Functions and Learning in ANNs
- Mcculloch & Pitz Neuron

2. Fully Connected Neural Networks

- Linear Perceptron, AdaLine and MadaLine
- Multi-Layer Perceptron (MLP)
- Auto-encoders and Restricted Boltzmann Machine
- Deep belief networks

3. Convolutional Neural Networks (CNNs)

- Convolutional Neural Networks (CNNs)
- Developments and Techniques in CNNs
- Some popular Architectures (AlexNet, VGG, ResNet, Inception, MobileNet, DenseNet, and EfficientNet)

4. Region based CNNs

- CNNs for Object Detection (RCNN, Mask RCNN, YOLO,...)
- CNNs for Object Segmentation (UNet, Mask RCNN, ...)

5. Recurrent Neural Networks

- Recurrent Neural Network(RNN)
- Long Short Term Memory (LSTM)
- Gated Recurrent Units(GRU)
- Some Extensions

6. Transformers

- Attention Mechanisms
- Vanilla Transformer
- Vision Based Transformers

7. Deep Generative Models

- Variational Auto-encoders
- Generative Adversarial Networks
- Flow-based Deep Generative Models (Invertible Transform of Distributions)
- Diffusion Models (Gradually add Gaussian Noise and then reverse)

Mini Projects and Exams

Chapters	2	3	4	5	6	7
Mini Projects 60%	**M. Pr. 1** 10%	**M. Pr. 2** 10%	**M. Pr. 3** 10%	**M. Pr. 4** 10%	**M. Pr. 5** 10%	**M. Pr. 6** 10%
Exams 40%	Midterm(Chapters1, 2, and 3) 20%			Final(Chapters4, 5, 6, and 7) 20%		
Total Score	100%					

*** A few (optional) mini-projects are designed for extra work (bonus points)**

Chapter 1

Introduction

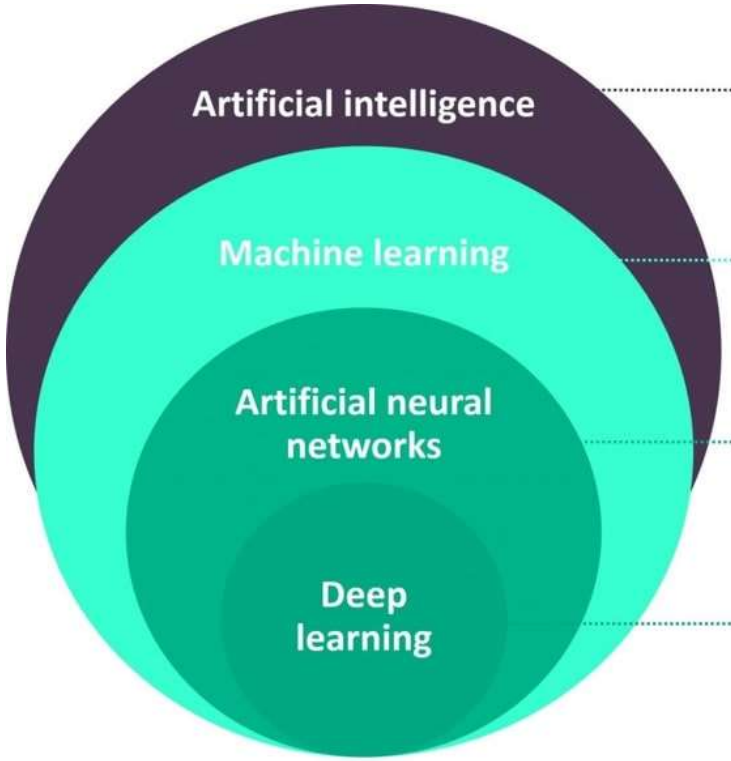
1. Introduction

What is Artificial intelligence?

- **AI is "a learning technology"**
- **Inspiring from the natural intelligence (in human and animals)**
  - **Environment perception (clustering/classification/regression)**
  - **Providing Memory about it**
  - **Interaction with it (Doing complexity tasks)**
  - **Computation intelligence (math-logic)**
- **Stablished on branches of mathematic:**
  - **Matrices Algebra in representation and in computation**
  - **Analysis in learning**
  - **Statistical and Geometric in interpretation and design**

- **Widely applicable in our learning activities**
  - **Media, Security and Entertainment**
  - **Cyber physical , hybrid and autonomous systems**
  - **Social Robots and human-robot interaction**
  - **Industrial systems**
  - **Health systems**
  - **Cognitive systems**
  - **Finance, and Energy Systems**

Artificial NN (ANN) and Deep Learning (DL) in AI



Artificial intelligence (AI)
Any techniques that enable machines to solve a task in a way like humans do

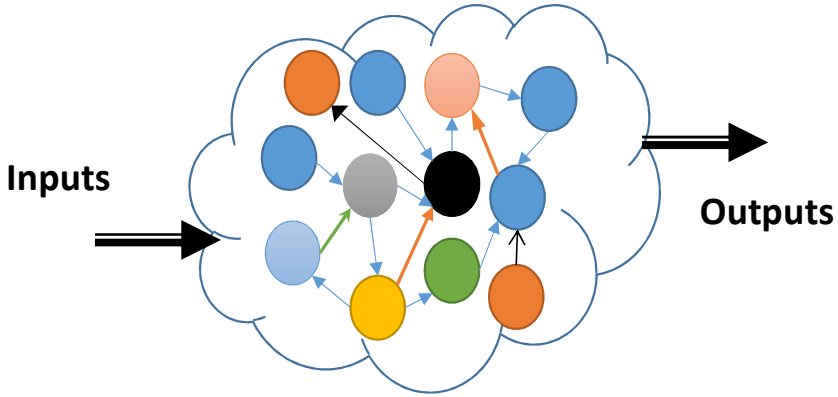
Machine learning (ML)
Algorithms that allow computers to learn from examples without being explicitly programmed

Artificial neural networks (ANN)
Brain-inspired machine learning models

Deep learning (DL)
A subset of ML which uses deep artificial neural networks as models and automatically builds a hierarchy of data representations

A simple definition for Neural Networks

- Neural Networks : A set of simple units (neurons) which are wired together in layers, in order to make important (desired) outputs against stimulating inputs.



Natural Neural Networks

- NNs have been created naturally in body of animals and plants.

Important questions about animals intelligence

- How do they recognize (cluster and classify) foods, enemies and environment?
- How do they make memories about different events?
- How do they learn to define different mechanisms and acts in their environment?

  ……..

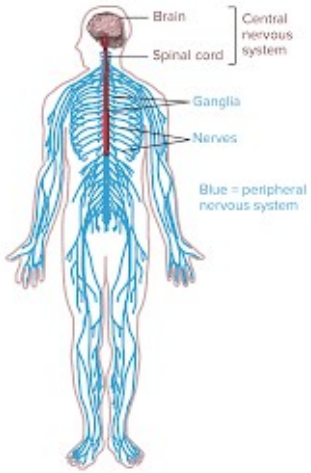
- **How do they become intelligent?**

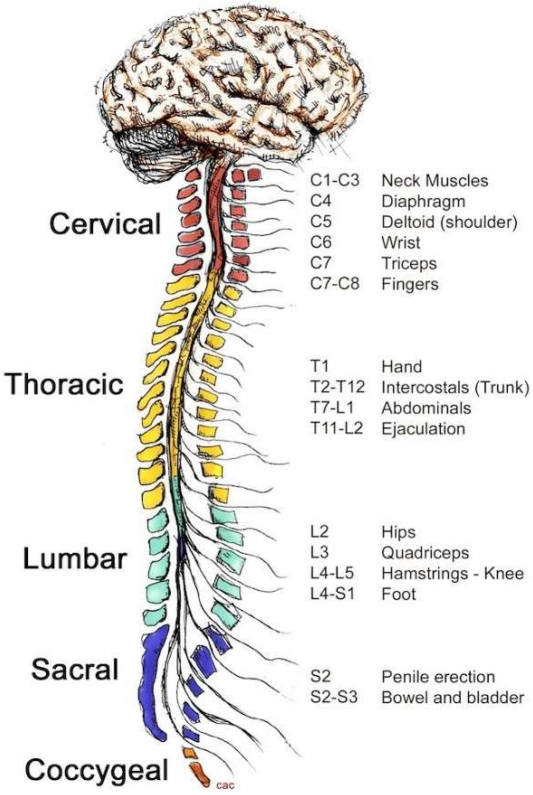
(Trying to answer to aforementioned questions)

Biologists and Scientists (about 200-300 years ago) discovered

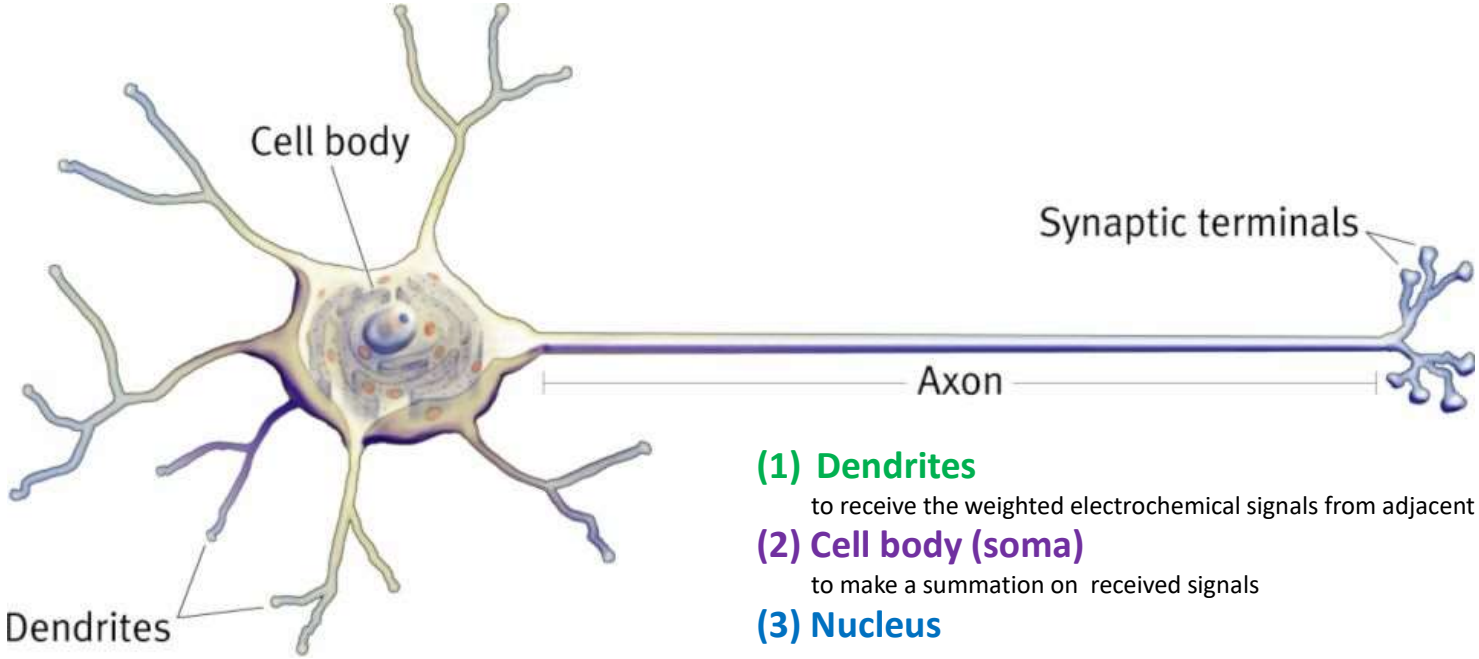
the brain, nervous system and spinal cord.

(currently) Neuroscientists research about the structure and function of the brain as the most important part of intelligent.





A Biological Neuron (nerve cell)



(1) Dendrites

to receive the weighted electrochemical signals from adjacent neurons

(2) Cell body (soma)

to make a summation on received signals

(3) Nucleus

to generate an impulsive signal by comparing the absorbed signal with a threshold

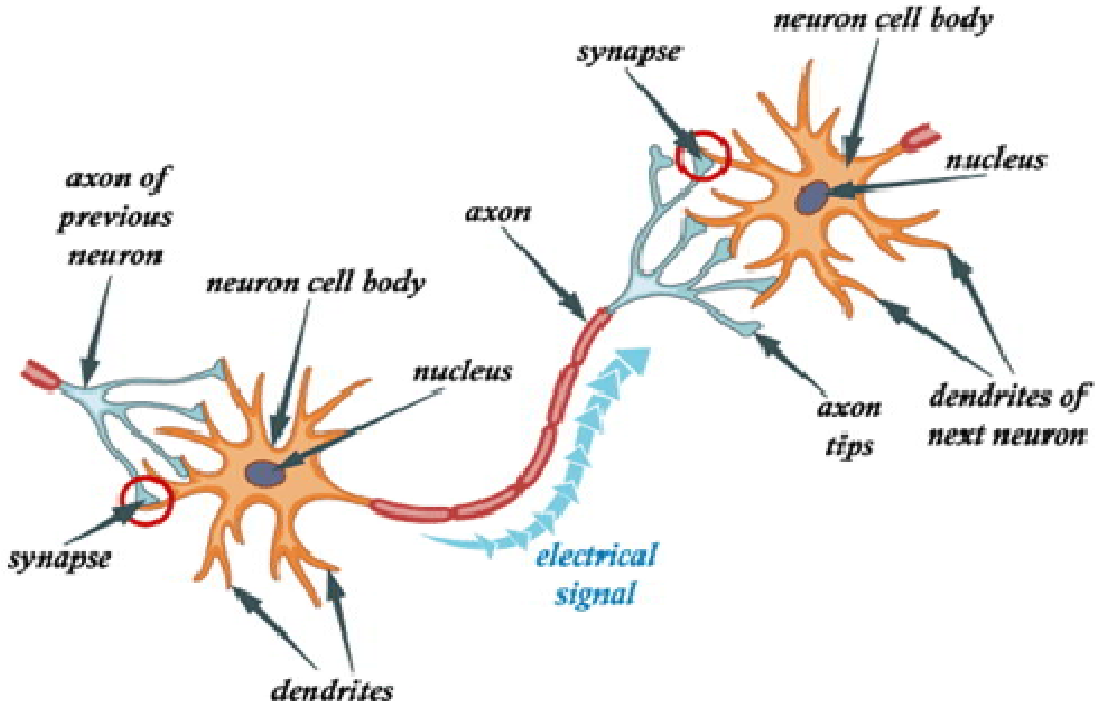
(4) Axon

to send the generated signal to other adjacent neurons

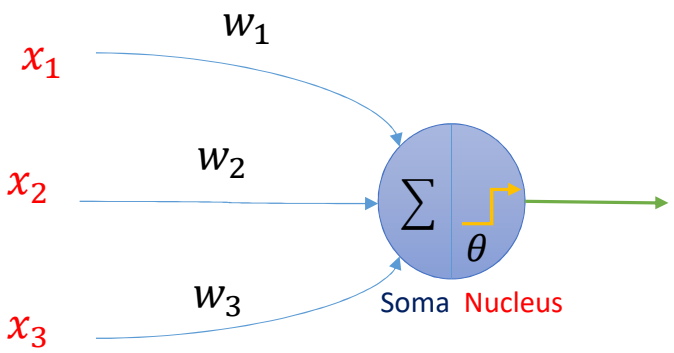
(5) Synaptic gaps

to assign a weight to each send signal to adjacent neurons

Communication between neurons by electrochemical signals:



A simple mathematical Model for Biological Neuron



$$z = w_1x_1+w_2x_2+w_3x_3$$

$$y = \begin{cases} 1 & z > \theta \\ 0 & otherwise \end{cases}$$

Dendrites Synaptic gaps *cell* *Axon*

Types of neurons in human nervous system

1. Sensory **neurons**

Get information about what's going on inside and outside of the body and bring that information into the central nervous system (CNS) so it can be processed.

2. Interneurons

which are found only in the CNS, connect one neuron to another. Most of interneurons are in the brain.

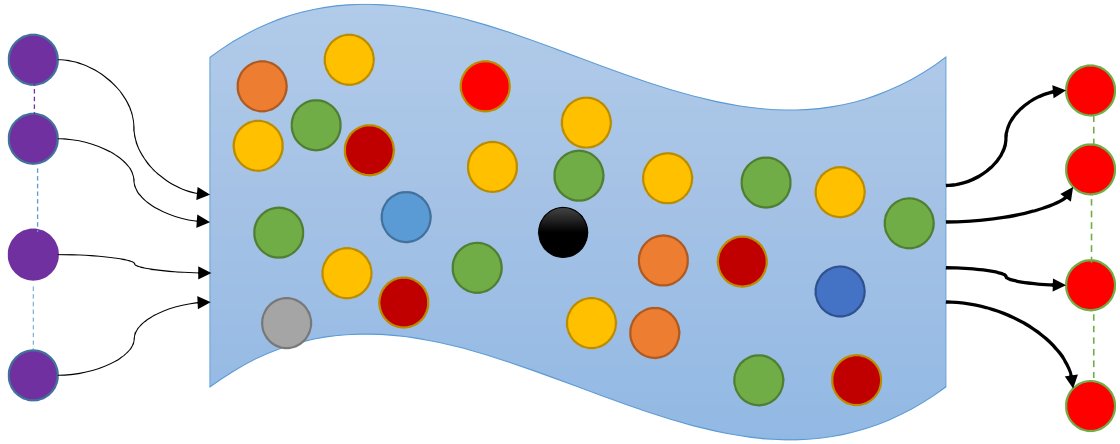
There are about 100 billiard neurons in the brain.

There are about 10^15 connections among neurons(10000 connections for each neuron on average)

3. Motor **neurons**

get information from other neurons and convey commands to your muscles, organs and glands.

A simple Input-Output Model for human nervous system



Sensory **neurons** **(Input Layer)**

Interneurons (**Hidden Layers**)

Motor **neurons** **(Output Layer)**

Learning capabilities of NNs in human body

(1) Classification

Localization, Detection and Classification of different objects, faces, voices, smells, and approximation and prediction different physical variables: distances, temperatures, smoothness, brightness, and so on…..

(2) Memory

Capability to create memories about different events with long and short dependencies.

memory about people, seasons, places, celebrations, …..

Capability to associate sequenced different patterns together.

(3) Complex and difficult tasks/actions

Car driving and parking, Swimming, Playing music,…..

(4) Computational intelligence

Logic, mathematics, Inference

Learning in natural neural network
(some important facts)

1. The learning process in brain is mainly performed by tuning the synaptic gaps. Information gathered or resulted from environment is coded in synaptic gaps.
2. The communication speed (electrochemical signal transition) for each neuron is low but Since the communications among neurons are performed in parallel, the processing speed of the brain is high on average.
3. The learning process in the brain is not disturbed if some parts of the brain are damaged and hence the robustness of the natural NNs is high (Fault tolerant).
4. Due to high level abstraction and inference in the brain, its generalization in learning is high.

Artificial Neural Networks

Inspiring From natural NNs:

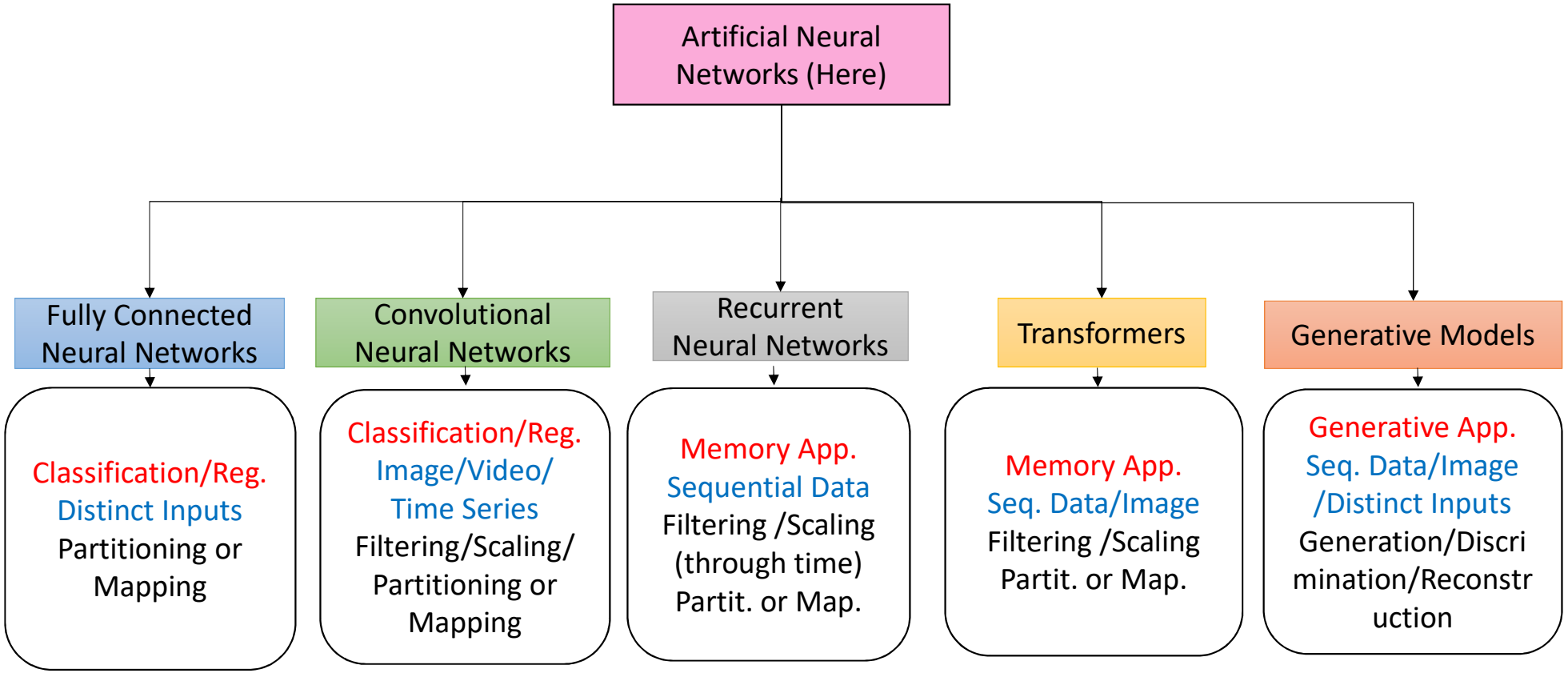
From begging of the 20- century, Scientists and engineers have been interested to design artificial neural networks:

"To make solutions for demanded (challenging) learning problems"

Applications of ANNs

- Classification NNs
  - Classification of objects, images, faces, texts, speeches, signatures, videos and ….
  - Disease Diagnosis
  - Fault Detection (Industrial)
  - Fraud/ Anomaly Detection
- Regression NNs
  - Function Approximation(Static/ODE/PDE)
  - Identification
  - Segmentation
  - Simulation
  - Prediction
  - Signal Recovering/ Repairing

- Memory NNs
  - Word/Voice/Video Prediction
  - Natural Language Translation
  - image captioning/Descriptions
  - Sentiment analysis
  - {Demand, Transaction and Price} Prediction
  - Path Planning, Tele Representation in Social robots
- Mechanism-Based NNs
  - Pattern Generation/sorting/clustering/Noise cancellation
  - Super-resolution techniques
  - Recommender systems
  - Image Enhancement
  - Data Compressing
  - Imitation-based systems
  - To describe the mechanism of the Brain



Sequential Data: Text, Time series, Speech, Video,....

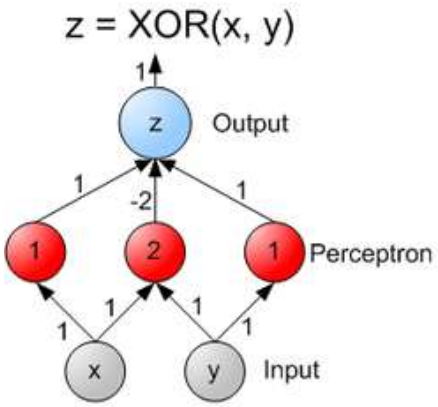
Feed-forward and recurrent neural networks

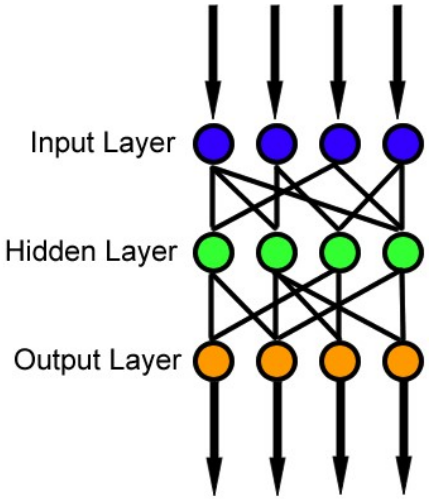
1- Feedforward Neural Networks

In feedforward networks, messages(data flow) are passed forward only

Feedforward neural network is an artificial neural network wherein connections between the nodes do not form a cycle.

Each output is a static of function of inputs.



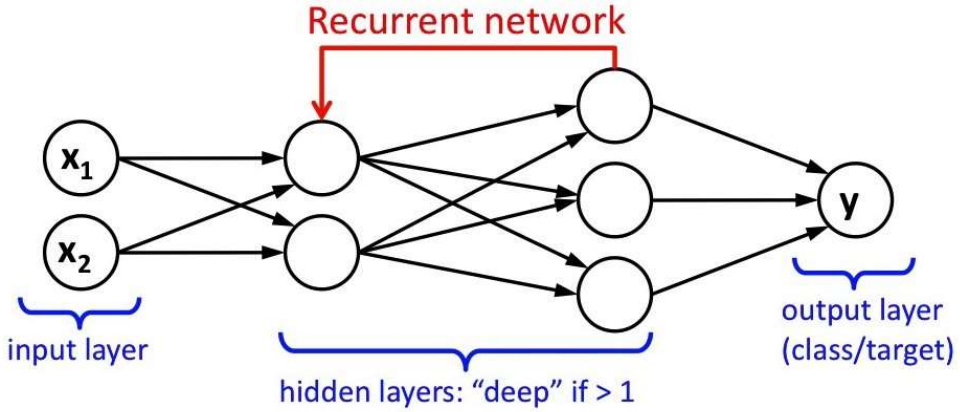


2- Recurrent Neural Networks

In recurrent networks, at least in one layer , messages(data flow) are returned to the same or former layers.

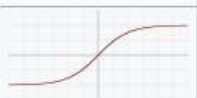
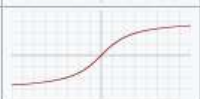
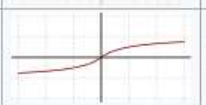
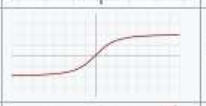
In recurrent neural networks (RNNs), a notion of time is introduced. The input at time step t depends on an output from time step t − 1.

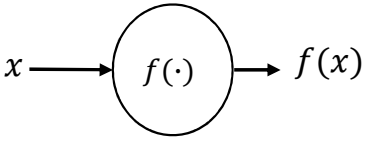
These networks are suitable to represent dynamic behavior of functions and systems as nonlinear difference or differential equations .



Activation Functions

- Each node(neuron) has an activation function by which it responses to the stimulating inputs.

Name	Plot	Equation	Derivative (with respect to x)	Range
Identity		$f(x) = x$	$f'(x) = 1$	$(-\infty, \infty)$
Binary step		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x \neq 0 \\ ? & \text{for } x = 0 \end{cases}$	$\{0, 1\}$
Logistic (a.k.a. Sigmoid or Soft step)		$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}}$ [1]	$f'(x) = f(x)(1 - f(x))$	$(0, 1)$
TanH		$f(x) = \tanh(x) = \frac{(e^x - e^{-x})}{(e^x + e^{-x})}$	$f'(x) = 1 - f(x)^2$	$(-1, 1)$
ArcTan		$f(x) = \tan^{-1}(x)$	$f'(x) = \frac{1}{x^2 + 1}$	$\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$
Softsign[7][8]		$f(x) = \frac{x}{1 + \lvert x \rvert}$	$f'(x) = \frac{1}{(1 + \lvert x \rvert)^2}$	$(-1, 1)$
Inverse square root unit (ISRU)[9]		$f(x) = \frac{x}{\sqrt{1 + \alpha x^2}}$	$f'(x) = \left(\frac{1}{\sqrt{1 + \alpha x^2}}\right)^3$	$\left(-\frac{1}{\sqrt{\alpha}}, \frac{1}{\sqrt{\alpha}}\right)$
Rectified linear unit (ReLU)[10]		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$[0, \infty)$



Supervised and unsupervised learning

1. Supervised Learning

Supervised learning as the name indicates a presence of supervisor as teacher. Basically **supervised learning** is a **learning** in which we teach or train the ANNs using data which is well labeled that means some data is already tagged with correct answer.

Applications: **Classification/Regression: Function Approximation and Prediction/Recognition**

Learning methods

- Error back propagation(Steepest Descends) MSE, MAE, Cross Entropy,…. : Batch based- stochastic point based-stochastic mini batch based

  Gradient (Local search): Descent optimization algorithms:

  SGD/SGD+Momentum/Nesterov accelerated gradient/Adagrad/RMSprop/Adam/AdaMax/Nadam/AMSGrad
- Evolutionary based (Global search) : **Genetic Algorithm, Particle Swarm Optimization, Ant colony**
- Intelligence-based (Global search): **Simplex- Simulated Annealing**

2. Unsupervised Learning

•Unsupervised learning is the training of ANNs using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance.

•Instead of explicit targets for the network, there are some statistical or geometric properties for the suitable output of the network.

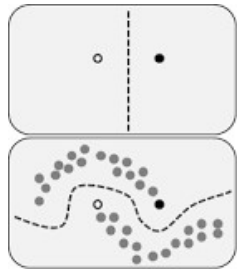
•Some examples of unsupervised learning algorithms include **K-Means Clustering, Principal Component Analysis and Hierarchical Clustering**.

•Causal Relationship in Regression Problems.

•Applications: **Pattern Generation/ Pattern clustering/ Pattern sorting/ Optimization problems/Control tasks**

3. Semi Supervised and Self Supervised Learning

- Semi-supervised learning is **an approach to machine learning that combines a small amount of labeled data with a large amount of unlabeled data during training**. Semi-supervised learning falls between unsupervised learning (with no labeled training data) and supervised learning (with only labeled training data).



- Self-supervised learning is **a representation learning method where a supervised task is created out of the unlabelled data**. Self-supervised learning is used to reduce the data labelling cost and leverage the unlabelled data pool. Some of the popular self-supervised tasks are based on contrastive learning.

What is the difference between unsupervised and self-supervised?

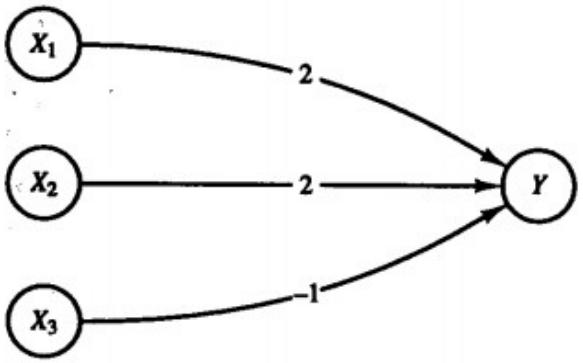
The only difference is that, unlike unsupervised learning, self-supervised learning does not perform the grouping and clustering of data, as is the case with unsupervised learning. This learning type allows machines to examine part of a data example to figure out the remaining part

Mcculloch & Pitz Neuron

Warren MuCulloch (neuroscientist) and Walter Pitts (logician) 1943

1. The activation of a McCulloch-Pitts neuron is binary. That is, at any time step, the neuron either fires (has an activation of 1) or does not fire (has an activation of 0).
2. McCulloch-Pitts neurons are connected by directed, weighted paths.
3. A connection path is excitatory if the weight on the path is positive; otherwise it is inhibitory. All excitatory connections into a particular neuron have the same weights.
4. Each neuron has a fixed threshold such that if the net input to the neuron is greater than the threshold, the neuron fires.
5. The threshold is set so that inhibition is absolute. That is, any nonzero inhibitory input will prevent the neuron from firing.
6. It takes one time step for a signal to pass over one connection link.

Figure 1.12 A simple McCulloch-Pitts neuron Y.



y_in = $2x_1 + 2x_2 - x_3$

$\theta: threshold$

$$f(y\_in) = \begin{cases} 1 & \text{if } y\_in \geq \theta \\ 0 & \text{if } y\_in < \theta \end{cases}$$

Some Logic Functions by M&P neurons

AND

x_1	x_2	→	y
1	1		1
1	0		0
0	1		0
0	0		0

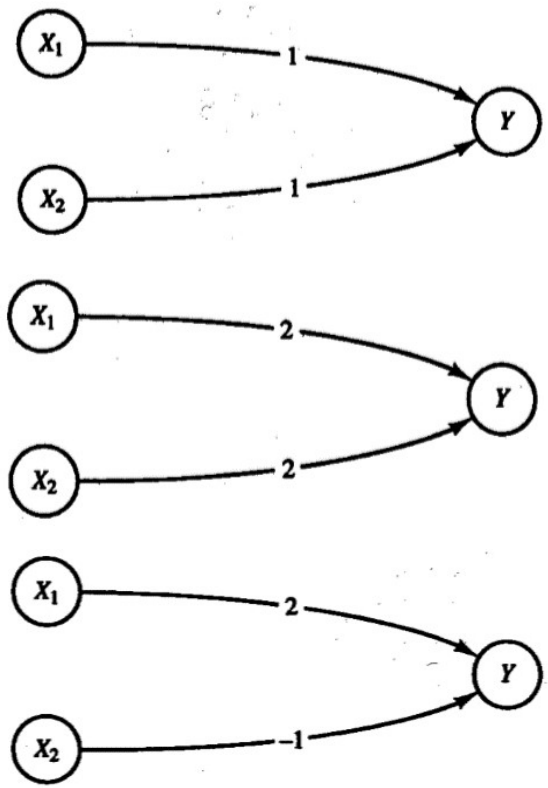
OR

x_1	x_2	→	y
1	1		1
1	0		1
0	1		1
0	0		0

AND Not

x_1	x_2	→	y
1	1		0
1	0		1
0	1		0
0	0		0

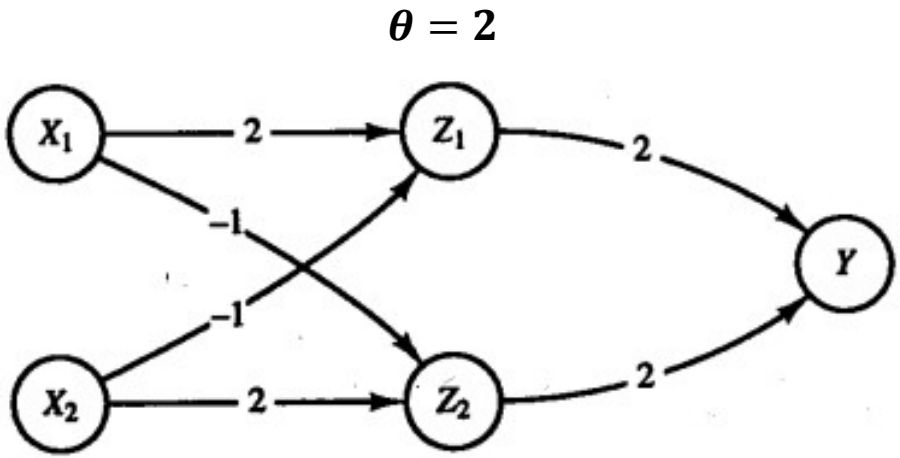
$\theta = 2$



Two Applications

- XOR

x_1	x_2	$\rightarrow$	y
1	1		0
1	0		1
0	1		1
0	0		0



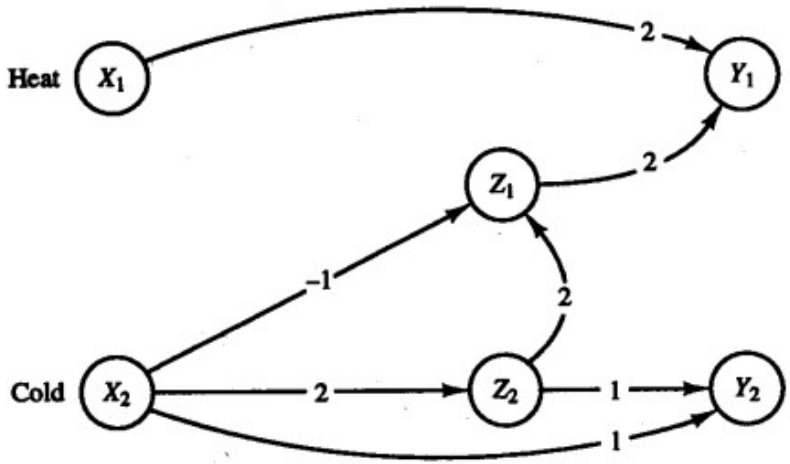
x_1 XOR $x_2 \leftrightarrow (x_1$ AND NOT $x_2)$ OR $(x_2$ AND NOT $x_1)$.

Two Applications

- Hot and Cold

Example 1.5 Modeling the Perception of Hot and Cold with a McCulloch-Pitts Net

It is a well-known and interesting physiological phenomenon that if a cold stimulus is applied to a person's skin for a very short period of time, the person will perceive heat. However, if the same stimulus is applied for a longer period, the person will perceive cold. The use of discrete time steps enables the network of McCulloch-Pitts neurons shown in Figure 1.18 to model this phenomenon. The example is an elaboration of one originally presented by McCulloch and Pitts [1943]. The model is designed to give only the first perception of heat or cold that is received by the perceptor units.



End of the Introduction

Thank you