Faculty of Sciences Tetouan
Mathematics Department
2023-2024

Introduction to Neural Networks: Basics

Prof. Youness EL YAZIDI

1 Motivations

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

Arthur Samuel (1959)

What is machine learning all about?

Machine Learning:

When we can use machine learning?

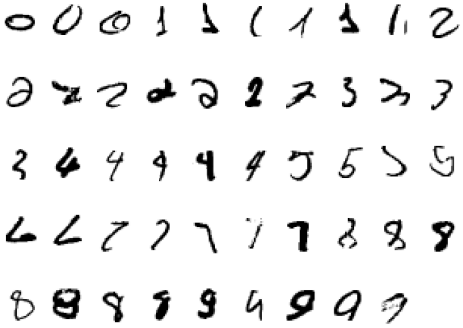
ML is used for example when:

- Human expertise does not exist (navigating on Mars)
- Humans can't explain their expertise (speech recognition)
- Models must be customized (personalized medicine)
- Models are based on huge amounts of data (genomics)

Can we use ML all the time!

No, there is no need to "learn" how to calculate payroll.

A classic example of a task that requires machine learning: It is very hard to say what makes a 1.



Types of Learning

- Supervised (inductive) learning:
  Given: training data + desired outputs (labels)
- Unsupervised learning:
  Given: training data (without desired outputs)
- Semi-supervised learning:
  Given: training data + a few desired outputs
- Reinforcement learning :
  Rewards from sequence of actions

How Supervised Learning work?

We are given n instances of input-output pairs

- **Inputs x** $= \{x_i\}$: covariates, predictors, features
- **Outputs y** $= \{y_i\}$: variates, targets, labels

The goal is: **Learn a function $f(\mathbf{x})$ to predict y given x.**

Neurophysiology

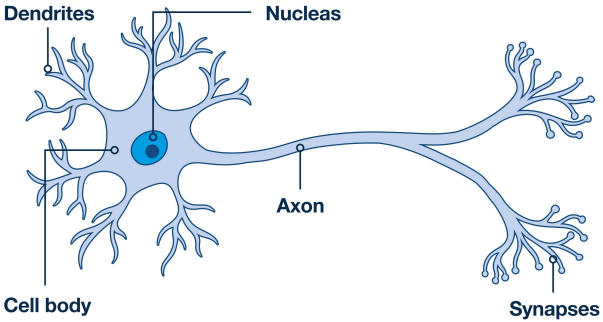
The human nervous system can be divided into three stages:

1. Receptors:
   - Convert stimuli from the external environment into electrical impulses
   - Rods and Cones of eyes,
   - Pain, touch, hot and cold receptors of skin.
2. Neural Net:
   - Receive information, process it and make appropriate decisions.
   - Brain
3. Effectors:
   - Convert electrical impulses generated by the the neural net (brain) into responses to the external environment.
   - Muscles and glands, speech generators.

Basic Components of Biological Neurons

The basic components of a biological neuron are:

- Cell Body (Soma) processes the incoming activations and converts them into output activations.
- Neuron Nucleus contains the genetic material (DNA).
- Dendrites form a fine filamentary bush each fiber thinner than an axon.
- Axon: Long thin cylinder carrying impulses from soma to other cells
- Synapses: The junctions that allow signal transmission b/w the axons and dendrites.



Dendrite: Receives signals from other neurons
Nucleus: Processes the information
Axon: Transmits the output of this neuron
Synapse: Point of connection to other neurons

Computation in Biological Neurons

- Incoming signals from synapses are summed up at the soma.
- On crossing a threshold, the cell fires generating an action potential in the axon hillock region.

What is a Neural network?

Neural network is a circuit or a **directed graph** (net of neurons), intended to solve a given optimization problems.

A neural network can be also defined as a nonlinear function, that receive inputs and link them with their appropriate outputs. It consists of basically:

1. Neurons: which pass input values through functions and output the result.
2. Weights: which carry values (real-number) between neurons.

Neurons can be categorized into layers:

1. Input Layer
2. Hidden Layer
3. Output Layer

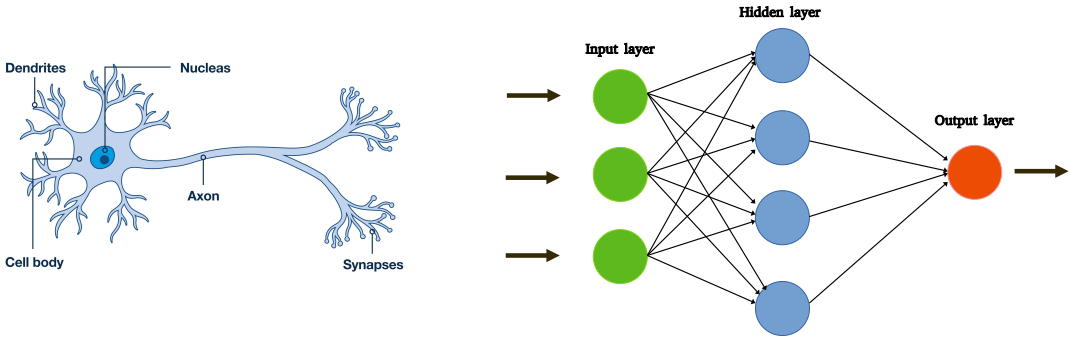


Figure 1: Architecture (Topology) of a Neural Network

Neural Network applications

NN is widely applied to various real-world data classification process, such as

1. Speech recognition
2. Product inspection
3. Fault detection
4. Bind rating
5. Handwriting recognition
6. Financial trend prediction
7. Optical character recognition
8. Face recognition
9. Image analysis

Examples of using NNs variants in real applications

Home selling Ad use info	Price Click	Real estate Online Advertising	Classic NNs
Image	Object	Tagging	Convolutional NNs
Audio Russian	Text English	Speech recognition Translation	Recurrent NNs
Radar info	Position of objects	Autonomous driving	Hybrid forms of NNs

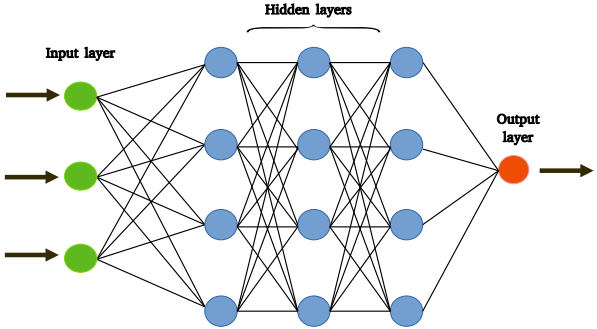


Figure 2: Architecture of a Convolutional Neural Network

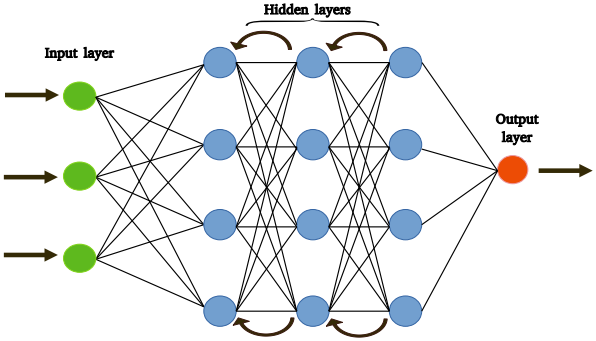


Figure 3: Architecture of a Recurrent Neural Network

NN is the algorithms mainly used for the purpose of learning and optimization based on the concept of the human brain. In general, the NN consists of the following five components:

1. The directed graph also called as network topology and whose arcs are referred as links.
2. Variable associated with the nodes.
3. Weight connected with the link.
4. Bias connected with the nodes.
5. The transfer function is used at each node that specifies the state of the node as the function based on its weight and bias. The transfer function either takes the step or sigmoid function to generate the output.

Neural network characteristics

The trained NN classifier has the following characteristics: Adaptive Learning, Mapping, Parallel operation and Prediction.

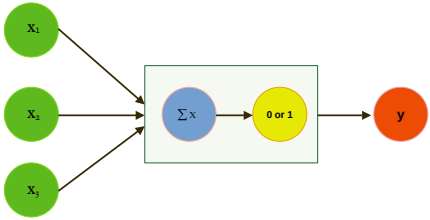
1. **Adaptive learning:** It has the facility to learn the performance of various tasks with respect to the training data.
2. **Self-organization(Mapping):** It creates its own representation or organization of data by receiving the proper information at learning time.
3. **Real-time operation:** It allows the hardware devices and parallel computation and takes the benefits of capability being manufactured and designed.
4. **Complex problem solving:** It has the capability to learn the faults and increases the performance by solving complex issues. Moreover, the major benefit of NN classifier that lies in the theoretical aspects is listed as follows:
5. **Predicting:** NNs have the capability to predict new outcomes.

Some basic Network Structures

1. The earliest NNs are The Perceptron, proposed by the psychologist Frank Rosenblatt (Psychological Review, 1958).
2. The Adaline (Adaptive Linear Neuron, due to B. Widrow, 1960). This artificial neuron is also known as the ALC (adaptive linear combiner), the ALC being its principal component. It is a single neuron, not a network.
3. The Madaline (Many Adaline), also due to Widrow (1988). This is an ANN (network) formulation based on the Adaline above, but is a multilayer NN.

McCulloch-Pitts Neuron

The first computational model of a neuron was proposed by Warren MuCulloch and Walter Pitts in 1943.



It is divided into 2 parts. The first one takes an input (dendrite) and performs an aggregation; and based on this aggregated value a decision is made in the second part.

The aggregation of inputs is performed with $g(x) = \sum x_i$; and the decision is

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

- θ is the threshold-step parameter
- g is the aggregation function
- f is the activation function

Example of making decision with M-P neuron

Make the decision watch a movie or not.
The result will be **0**: watch it; **1**: not watch it.

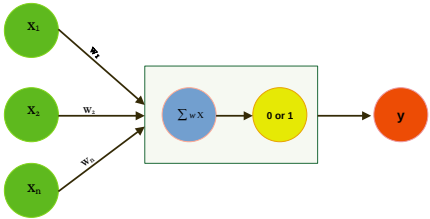
x_1: is it long? I shouldn't sleep late.
x_2: is it an horror type? I don't want to get nightmare.
x_3: is it English? I rather to not watch something French.
x_3: is it new released or old school movie?

For example: $x_1 = 1$, $x_2 = 0$, $x_3 = 0$ and $x_4 = 1$ then $g(x) = \sum x_i = 2$
We assume that $\theta = 2$, thus $g(x) \geq \theta$ thereafter the decision is to **watch** this movie.

Limitations McCulloch-Pitts Neuron

1. It handles only binary data
2. Can not incorporate weighted inputs
3. Generate binary output.

It was proposed by Rosenblatt in 1958. It was introduced to replace McCulloch-Pitts Neuron. It is a single layer neural network whose weights and biases could be trained to produce a correct target vector when presented with the corresponding input vector.



The aggregation of inputs is performed with $g(x) = \sum w_i x_i$; and the decision is

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

Here w_i is a bias or weight to control the importance of each value x_i.

Example of making decision with Perceptron

Make the decision watch a movie or not.
The result will be **0**: watch it; **1**: not watch it.

x_1: is it long? I shouldn't sleep late.	$w_1 = 0.5$
x_2: is it an horror type? I don't want to get nightmare.	$w_2 = 1$
x_3: is it English? I rather to not watch something French.	$w_3 = -0.5$
x_3: is it new released or old school movie?	$w_4 = 0$

For example: $x_1 = 1$, $x_2 = 0$, $x_3 = 0$ and $x_4 = 1$ then $g(x) = \sum w_i x_i = 0.5$
Again, we take $\theta = 2$, thus $g(x) < \theta$ thereafter the decision is to **not watch** this movie.

Limitations of perceptron

1. It handles only binary data
2. Use a linear activation function
3. Generate binary output.

ToDo

A small research on learning using Adaline.