TD 7 – Learning I Supervised Learning

Practical Information

TD Assistant
eponiatowski@clipper.ens.psl.eu

TD Material
https://github.com/esther-poniatowski/2223_UlmM2_ThNeuro

Goals of the TD

Three types of learning algorithms will be studied in this series of TDs : supervised, unsupervised and reinforcement learning.
This TD presents the standard paradigm of the **supervised perceptron**. The perceptron model originates with Rosenblatt in 1958 and marks the beginning of Artificial Neural Networks, which have in recent years attained unexpectedly high performance in various tasks, placing learning as a central question of Neuroscience.

Part **1** *focuses on*

Part **2** *considers*

1 Perceptron model

Perceptron

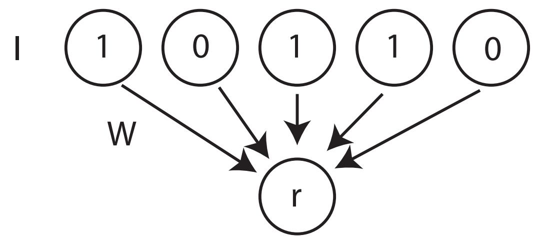
The **Perceptron** can be thought of as a single 'output' neuron receiving signals from N 'input' neurons.

Neurons are binary units, they can either be active or not. The input received by the output neuron at any time can be represented by a vector $\vec{I}$ of 0 and 1, in which the entry I_n is the activity of neuron n.

Inputs neurons communicate with the output neuron through synapses. Synaptic strengths are contained in a vector of weights $\vec{W}$, in which the entry W_n is the strength of the synapse connecting neuron n to the output neuron.

The output neuron is active if the total input it receives is larger than a threshold θ.

The function of the output neuron is to associate a set of P input patterns $\vec{I_p}$ with a desired output r_p. Learning consists in adjusting the parameters $\vec{W}$ and θ so as to solve this task.



① Give an expression of the total input received by the output neuron when a given pattern $\vec{I_p}$ is presented.

② Write a condition on the input patterns $\vec{I_p}$, the desired outputs r_p and the parameters $\vec{W}$ and θ such that the task is solved.

For convenience, consider an imaginary input $I_0 = 1$ which is always turned on.

③ Rewrite the condition as a function of $(I_0, \vec{I_p})$ and $(W_0, \vec{W})$ where $W_0 = -\theta$.

④ Find a set of input patterns $\vec{J_p}$ such that the condition is equivalent to :

$$\forall p, \quad \vec{J_p} \cdot \vec{W} > 0$$

⑤ Explain the meaning of the ability of a network to *generalize* after learning.

2 Perceptron algorithm

Perceptron algorithm

To correctly associate inputs to their respective output, the perceptron is trained according to the following learning algorithm :

- Randomly pick an input pattern $\vec{J}_p$
- If $\vec{J}_p \cdot \vec{W} > 0$, pick a new input pattern.
- If $\vec{J}_p \cdot \vec{W} < 0$, perform a learning step : $\vec{W}(t+1) = \vec{W}(t) + \epsilon \vec{J}_p$

The goal of this part is to show that if there exists a solution $\vec{W}^*$, then this algorithm necessarily finds a solution.

For this, the following notation will be used for the *angle* between two vectors :

$$\cos[\alpha(t)] = \frac{\vec{W}(t) \cdot \vec{W}^*}{\|\vec{W}(t)\| \|\vec{W}^*\|} \tag{1}$$

with $\alpha(t)$ the angle between $\vec{W}(t)$ and $\vec{W}^*$.

6 Explain why this algorithm is an implementation of "supervised" learning.

7 Introduce $l = \min_p \vec{J}_p \cdot \vec{W}^* > 0$ and find a lower bound on $\vec{W}(t+1) \cdot \vec{W}^*$ given $\vec{W}(t) \cdot \vec{W}^*$. Considering $\vec{W}(0) = 0$ deduce a lower bound on $\vec{W}(t) \cdot \vec{W}^*$.

8 Introduce $L = \max_p \|\vec{J}_p\|^2$ and find an upper bound on $\|\vec{W}(t+1)\|^2$ given $\|\vec{W}(t)\|^2$. Deduce an upper bound on $\|\vec{W}(t)\|^2$.

9 Find a lower bound on $\cos[\alpha(t)]$.

$$\cos[\alpha(t)] \geq \frac{t\epsilon l}{\sqrt{t\epsilon^2 L}} = \sqrt{t}\frac{l}{L}$$

10 Explain why the algorithm necessarily finds a solution.