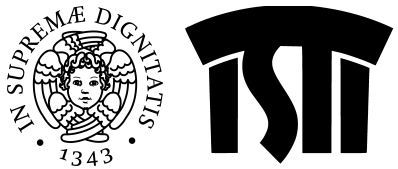
# ARTIFICIAL INTELLIGENCE 2

## The Perceptron

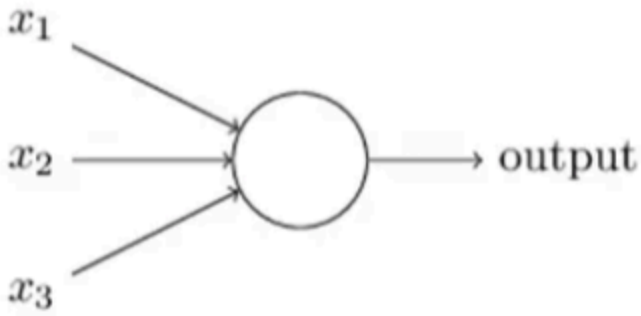
*Francesco Spinnato*

* francesco.spinnato@unipi.it
University of Pisa



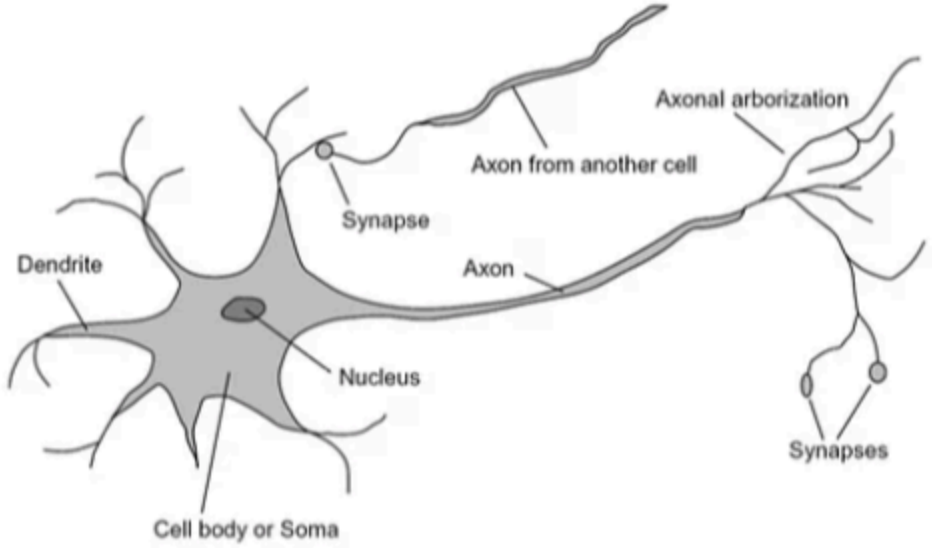
# A Biological Inspiration

Artificial neural networks are loosely inspired by the **human brain**.




Each **biological neuron** in the brain operates through electrochemical signals and consists of:

- a **cell body** with **dendrites** receiving inputs;
- an **axon** sending outputs.

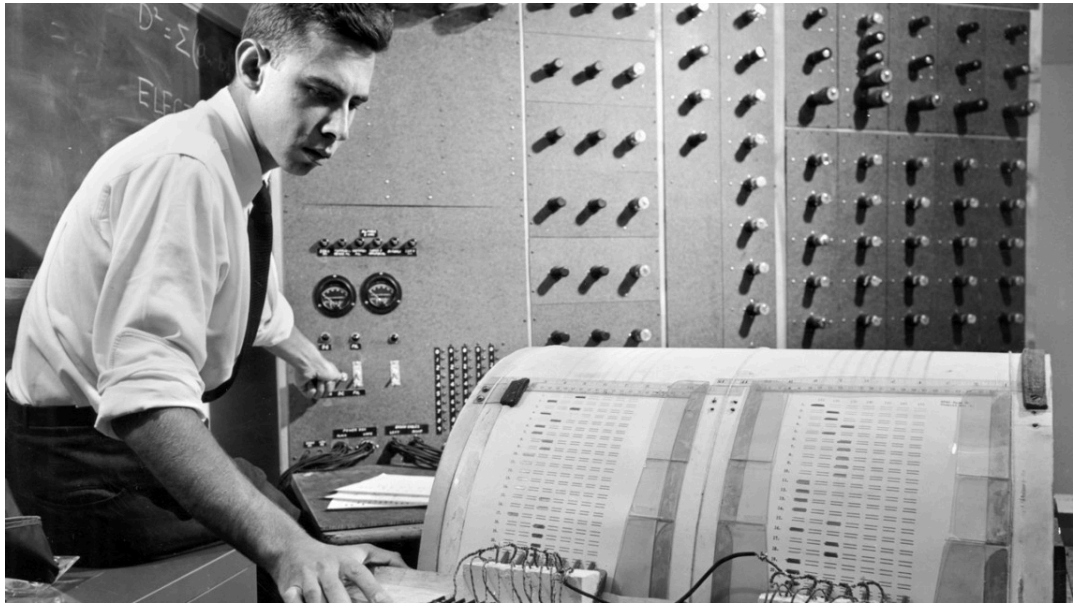



https://www.pearsonhighered.com/assets/samplechapter/0/1/3/1/0131471392.pdf (1.1, 1.2, Table 1.1)

# Computational Model of a Biological Neuron

The **artificial neuron network** was invented in 1943 by Warren McCulloch and Walter Pitts in "A logical calculus of the ideas immanent in nervous activity."

The **perceptron** was first simulated by Frank Rosenblatt in 1957 on an IBM 704.



# From Linear Regression to the Perceptron

We have a dataset with $M$ features and $N$ rows, and a target column.

| | $1$ | $2$ | $3$ | $\dots$ | $M$ | $y$ |
|---|---|---|---|---|---|---|
| $[1]$ | $x_1^{[1]}$ | $x_2^{[1]}$ | $x_3^{[1]}$ | $\dots$ | $x_M^{[1]}$ | $y^{[1]}$ |
| $[2]$ | $x_1^{[2]}$ | $x_2^{[2]}$ | $x_3^{[2]}$ | $\dots$ | $x_M^{[2]}$ | $y^{[2]}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ | $\vdots$ |
| $[N]$ | $x_1^{[N]}$ | $x_2^{[N]}$ | $x_3^{[N]}$ | $\dots$ | $x_M^{[N]}$ | $y^{[N]}$ |

The multiple linear regression model can be written as:

$$\hat{y}^{[i]} = w_1 x_1^{[i]} + w_2 x_2^{[i]} + \cdots + w_M x_M^{[i]} + b$$

# From Linear Regression to the Perceptron

Let's just focus on a single row $i$

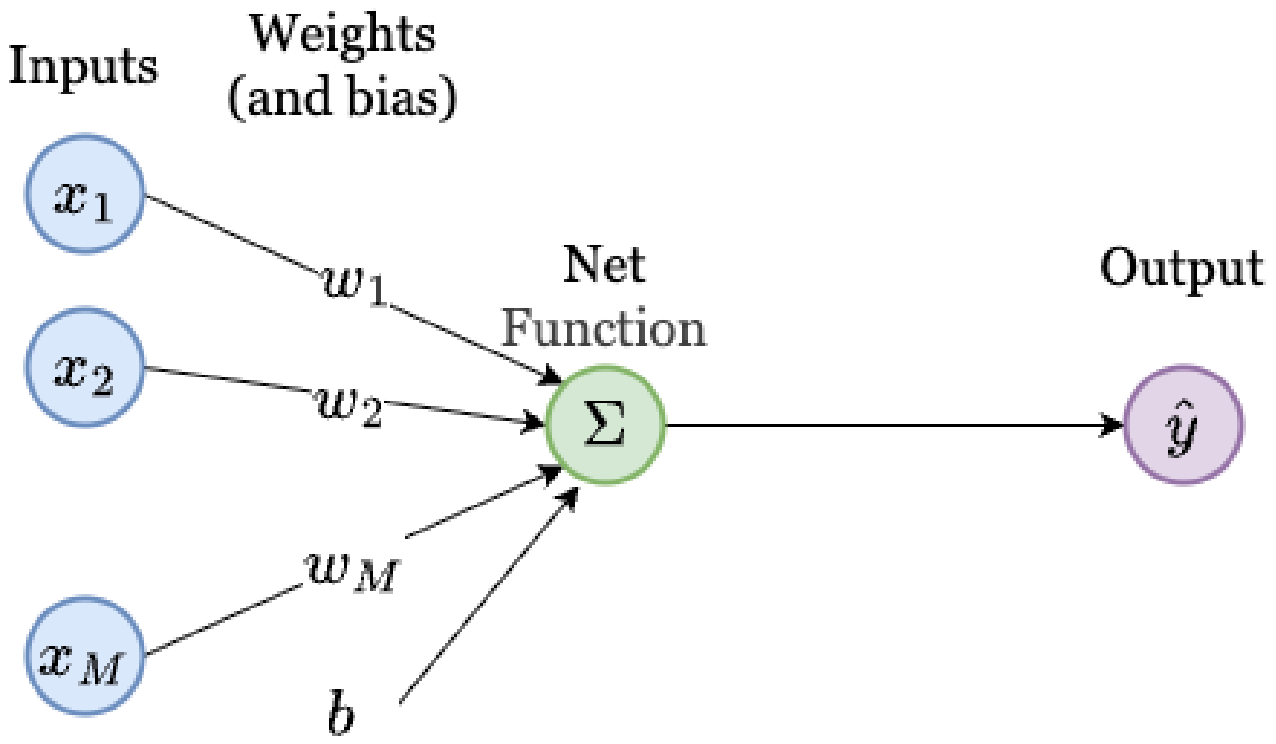
| | $1$ | $2$ | $3$ | $\dots$ | $M$ | $y$ |
|---|---|---|---|---|---|---|
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $[i]$ | $x_1^{[i]}$ | $x_2^{[i]}$ | $x_3^{[i]}$ | $\dots$ | $x_M^{[i]}$ | $y^{[i]}$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |

An drop the instance index $i$ for easier notation:

$$\hat{y} = w_1 x_1 + w_2 x_2 + \cdots + w_M x_M + b$$

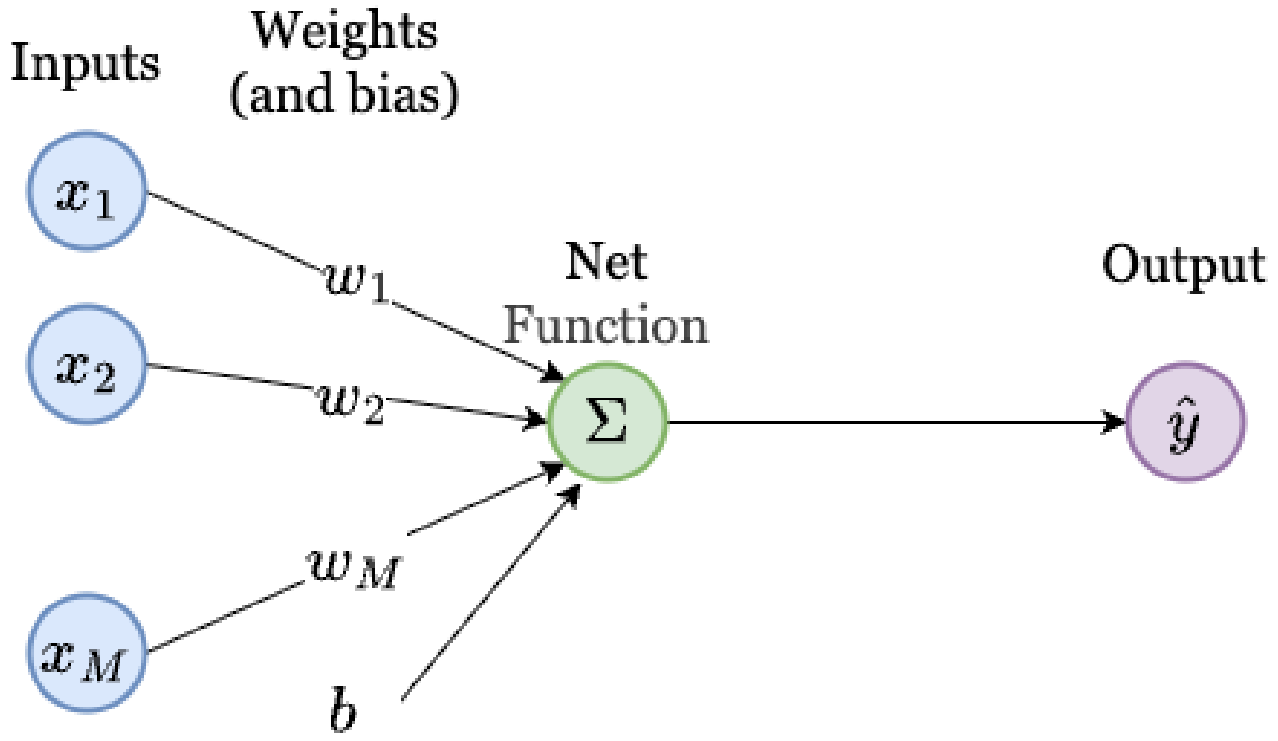
# From Linear Regression to the Perceptron

$$\hat{y} = w_1 x_1 + w_2 x_2 + \cdots + w_M x_M + b$$



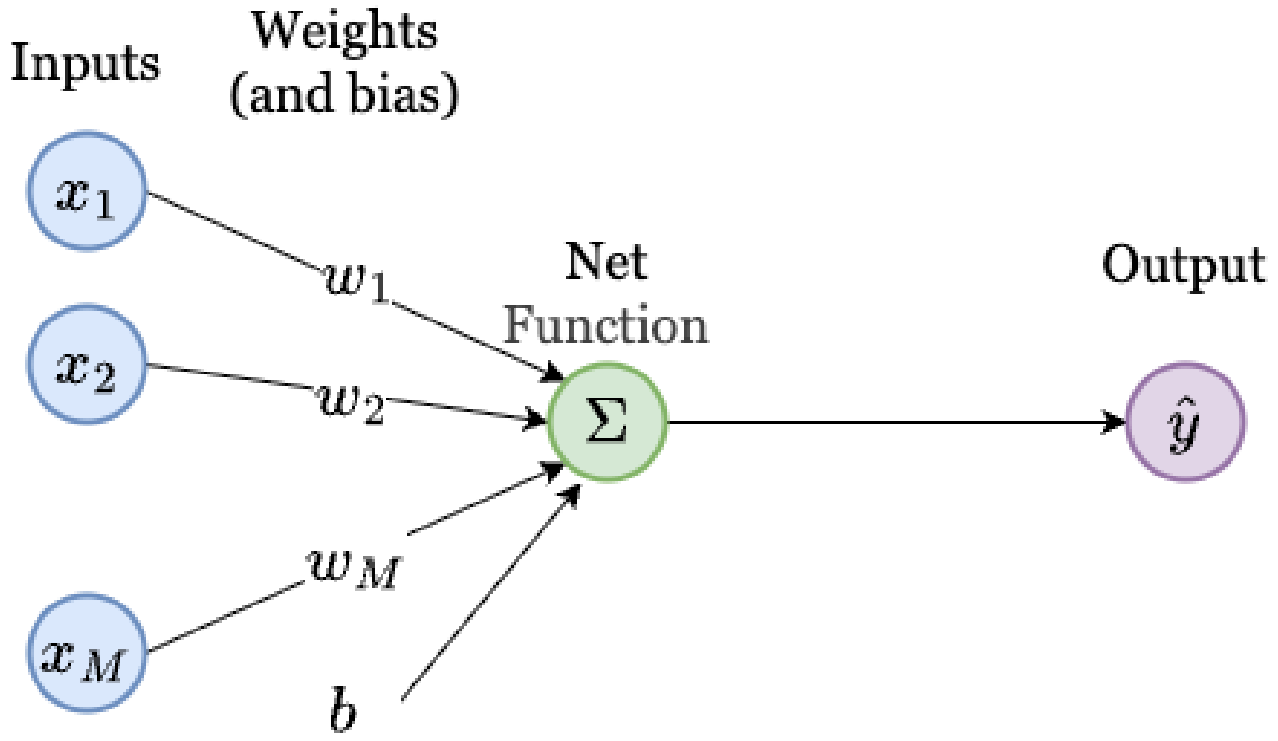
# From Linear Regression to the Perceptron

$$\hat{y} = \sum_{j=1}^{M} w_j x_j + b$$



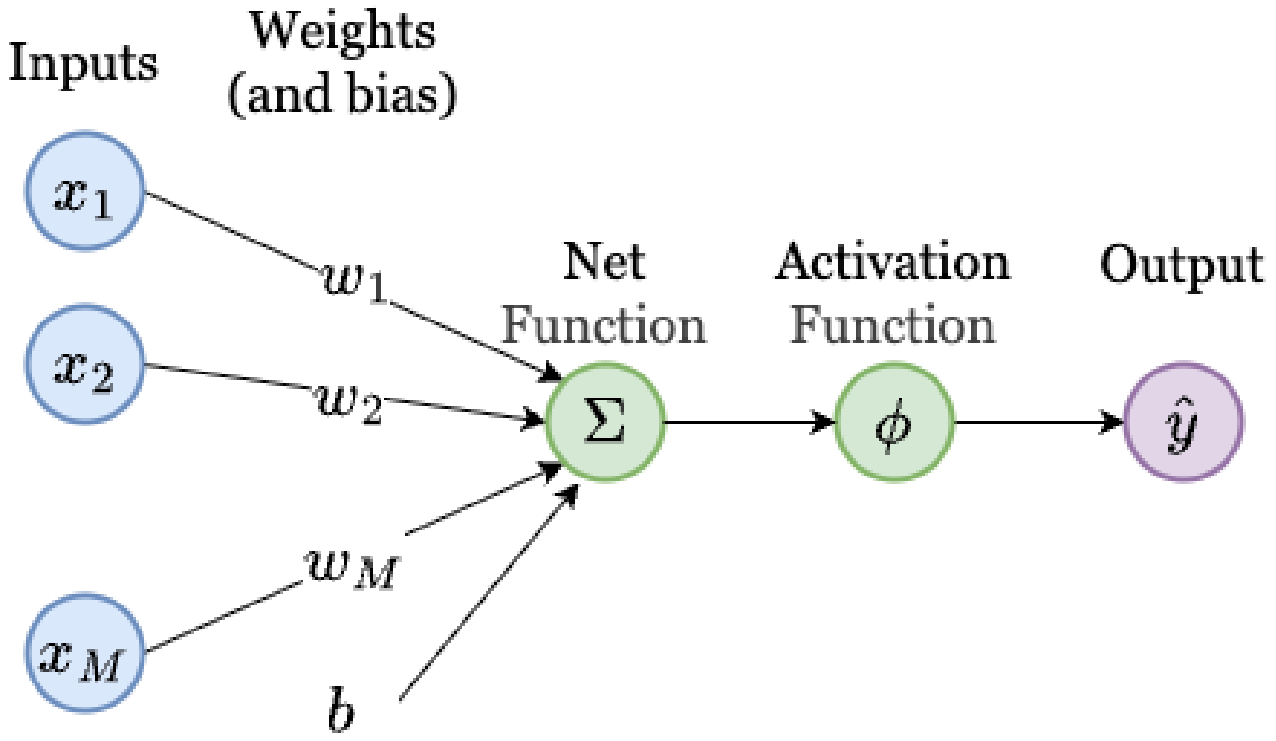
# From Linear Regression to the Perceptron

$$\hat{y} = \sum_{j=1}^{M} w_j x_j + b$$

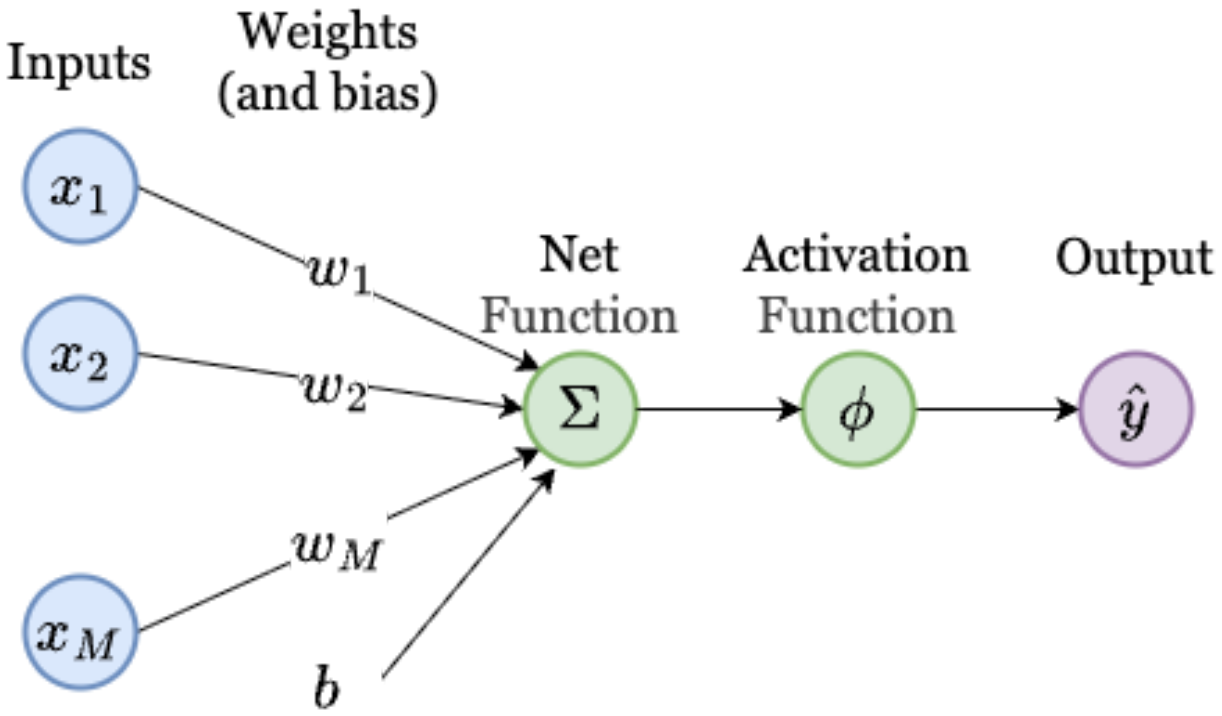


# The Perceptron

$$\hat{y} = \phi\left(\sum_{j=1}^{M} w_j x_j + b\right) = \phi(z)$$



# The Perceptron




$\phi$ is called the **activation function** and in the classical perceptron it is the **sign** function:

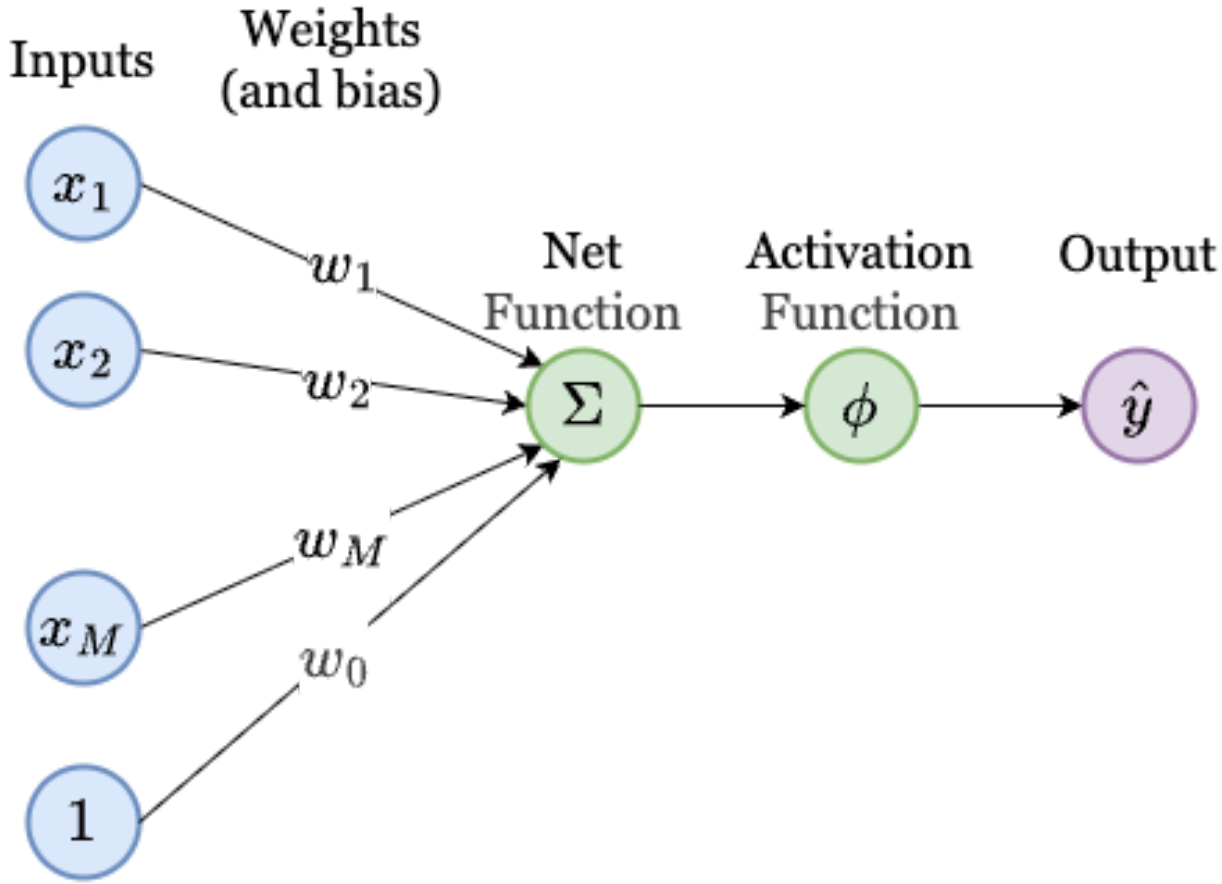
$$\hat{y} = \text{sign}(z) = \begin{cases} 1 & \text{if} \quad z \geq 0 \\ -1 & \text{if} \quad z < 0 \end{cases}$$

$$\boxed{\hat{y} = \text{sign}\left(\sum_{j=1}^{M} w_j x_j + b\right)} = \text{sign}(z)$$

# The Perceptron (alternative formulation)

For ease of notation sometimes the bias is replaced by an additional input neuron $x_0 = 1$, multiplied by a weight $w_0$.

$$\hat{y} = \text{sign}\left(\sum_{\boxed{\text{j=0}}}^{M} w_j x_j\right) = \text{sign}(z)$$



# Perceptron Inference Example

| $x_1$ | $x_2$ | $y$ |
|---|---|---|
| 3 | 1 | $-1$ |

Let the perceptron parameters be:

$$w_1 = 1, \quad w_2 = -2, \quad b = 0.5$$

For the input $(x_1, x_2) = (3, 1)$:

$$z = w_1 x_1 + w_2 x_2 + b = 1 \cdot 3 + (-2) \cdot 1 + 0.5 = 1.5$$

$$\hat{y} = \text{sign}(z) = 1 \quad \text{wrong prediction!}$$

# The Perceptron Learning Algorithm

Note that the, contrary to linear regression, the output of a classical perceptron is a **label**, either positive $1$, or negative $-1$.

So the perceptron is a (binary) **classification** model!

We want to find the best weights to solve the classification task so we need to train (modify) the weights when the model is wrong.

To train a perceptron we use an **online** procedure, i.e., we update the weights one example at a time, immediately after each misclassified sample.

# The Perceptron Learning Algorithm

1. Initialize the weights $[w_0, \ldots, w_M]$ (usually at $\mathbf{0}$);

**Repeat:** for each training example $\mathbf{x}^{[i]} = [x_1^{[i]}, \ldots, x_M^{[i]}]$, with label $y^{[i]}$:

2. Inference: compute

$$\hat{y}^{[i]} = f(\mathbf{x}^{[i]}) = \text{sign}\left(\sum_{j=0}^{M} w_j x_j^{[i]}\right);$$

3. Update the weights with (only if $\hat{y}^{[i]} \neq y^{[i]}$):

$$\mathbf{w}_{\text{new}} = \mathbf{w}_{\text{old}} + \underbrace{\lambda}_{\substack{\text{learning} \\ \text{rate}}} \underbrace{(y^{[i]} - \hat{y}^{[i]})}_{\text{error}} \mathbf{x}^{[i]}$$

**until one full pass over the data (epoch) produces no change in the weights.**

# The Perceptron Learning Algorithm

$$\mathbf{w}_{\text{new}} = \mathbf{w}_{\text{old}} + \underbrace{\lambda}_{\substack{\text{learning} \\ \text{rate}}} \underbrace{(y^{[i]} - \hat{y}^{[i]})}_{\text{error}} \mathbf{x}^{[i]}$$

$e = (y^{[i]} - \hat{y}^{[i]})$ is the error, i.e., the difference between the real label and the predicted one. It can be:

- $0$ when $y^{[i]} = \hat{y}^{[i]}$ so no update is needed
- $2$ if $y^{[i]} = 1$ and $\hat{y}^{[i]} = -1$: weight must be increased so that $f(\mathbf{x})$ will increase
- $-2$ if $y^{[i]} = -1$ and $\hat{y}^{[i]} = 1$: weight must be decreased so that $f(\mathbf{x})$ will decrease

$\lambda \in (0, 1]$ is the learning rate used to control the amount of adjustment made in each iteration.

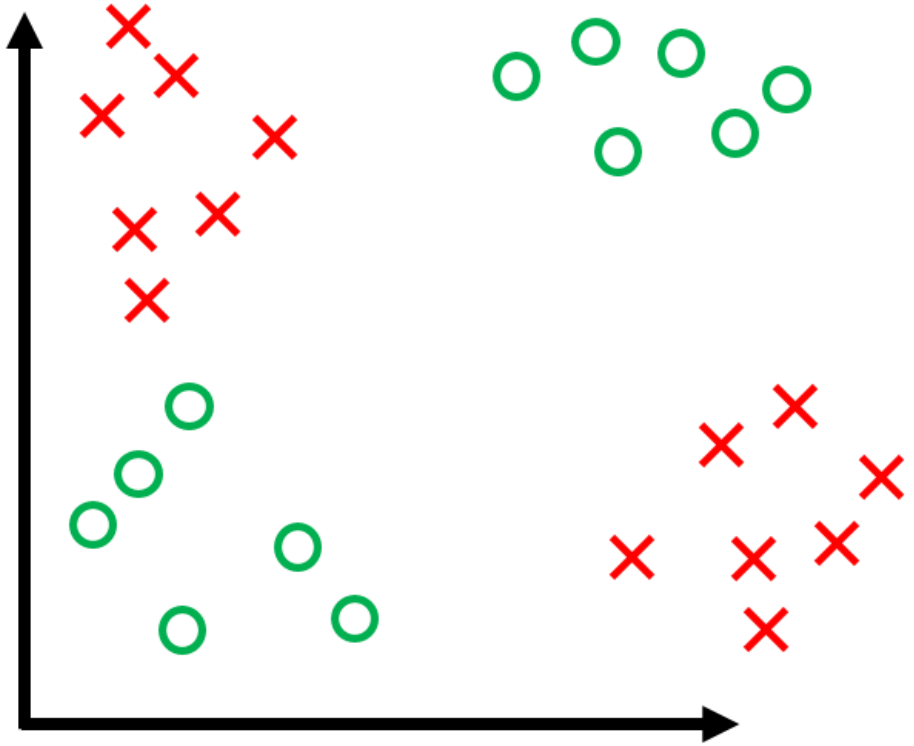
---

An equivalent formulation of the perceptron update is often presented using only the true label in the adjustment step. With labels encoded as −1 and +1, both forms are identical up to a constant factor, which can be absorbed into the learning rate.

# Perceptron Convergence

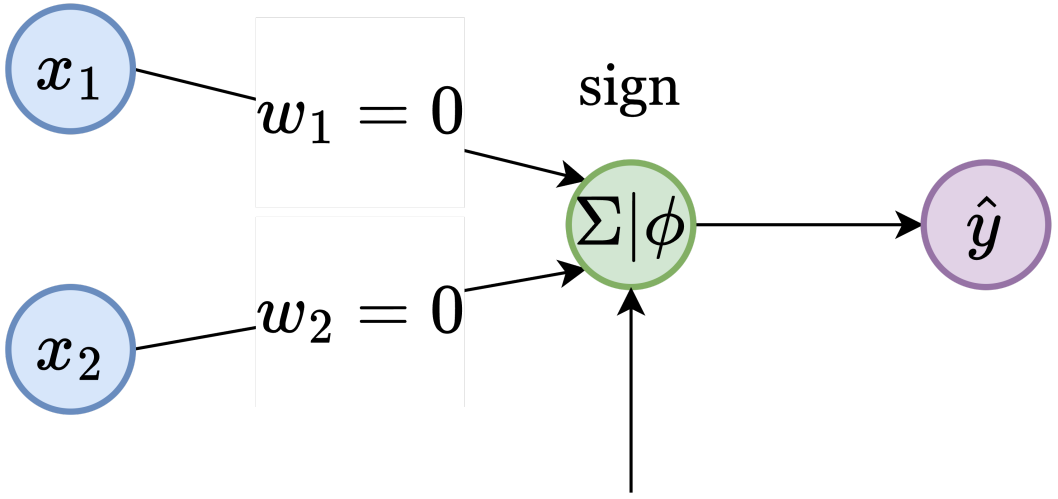
The Perceptron was arguably one of the first algorithm with a strong formal guarantee.

- If a dataset is **linearly separable**, the Perceptron will find a separating hyperplane in a finite number of updates.
- If the data is **not linearly separable**, it will loop forever.



# Example

| | $x_1$ | $x_2$ | $y$ |
|---|---|---|---|
| [1] | 0 | 1 | $-1$ |
| [2] | 0 | 0 | 1 |
| [3] | 1 | 2 | $-1$ |




$$\lambda = 0.3$$

$$\phi = \text{sign}$$

# Example - Iteration 1.

$$\mathbf{x}^{[1]} = [0, 1], \quad y^{[1]} = -1$$

$$\hat{y} = \text{sign}\Big(\underbrace{w_1}_{0} \cdot \underbrace{x_1}_{0} + \underbrace{w_2}_{0} \cdot \underbrace{x_2}_{1} + \underbrace{w_0}_{-1}\Big)$$

$$= \text{sign}(-1) = -1$$

$$e = \underbrace{y}_{-1} - \underbrace{\hat{y}}_{-1} = 0 \qquad \text{OK}$$

| | $w_0$ | $w_1$ | $w_2$ | $e$ |
|---|---|---|---|---|
| $[1]$ | $-1$ | $0$ | $0$ | $0$ |

# Example - Iteration 2.

$$\mathbf{x}^{[2]} = [0, 0], \quad y^{[2]} = 1$$

$$\begin{aligned}\hat{y} &= \text{sign}(0 \cdot 0 + 0 \cdot 0 - 1) \\ &= \text{sign}(-1) = -1\end{aligned}$$

$$e = 1 - (-1) = 2 \qquad \text{UPDATE}$$

$$w_1 \leftarrow \underbrace{w_1}_{0} + \underbrace{\lambda}_{0.3} \cdot \underbrace{e}_{2} \cdot \underbrace{x_1}_{0} = 0$$

$$w_2 \leftarrow 0 + 0.3 \cdot 2 \cdot 0 = 0$$

$$w_0 \leftarrow -1 + 0.3 \cdot 2 \cdot \underbrace{1}_{x_0=1} = -0.4$$

| | $w_0$ | $w_1$ | $w_2$ | $e$ |
|---|---|---|---|---|
| $[1]$ | $-1$ | $0$ | $0$ | $0$ |
| $[2]$ | $-1$ | $0$ | $0$ | $2$ |

# Example - Iteration 3.

$$\mathbf{x}^{[3]} = [1, 2], \quad y^{[3]} = -1$$

$$\begin{aligned} \hat{y} &= \text{sign}(0 \cdot 1 + 0 \cdot 2 - 0.4) \\ &= \text{sign}(-0.4) = -1 \end{aligned}$$

$$e = -1 - (-1) = 0 \qquad \text{OK}$$

| | $w_0$ | $w_1$ | $w_2$ | $e$ |
|---|---|---|---|---|
| $[1]$ | $-1$ | $0$ | $0$ | $0$ |
| $[2]$ | $-1$ | $0$ | $0$ | $2$ |
| $[3]$ | $-0.4$ | $0$ | $0$ | $0$ |

# Example - Iteration 4.

$$\mathbf{x}^{[1]} = [0, 1], \quad y^{[1]} = -1$$

$$\begin{aligned} \hat{y} &= \text{sign}(0 \cdot 0 + 0 \cdot 1 - 0.4) \\ &= \text{sign}(-0.4) = -1 \end{aligned}$$

$$e = -1 - (-1) = 0 \qquad \text{OK}$$

| | $w_0$ | $w_1$ | $w_2$ | $e$ |
|---|---|---|---|---|
| $[1]$ | $-1$ | $0$ | $0$ | $0$ |
| $[2]$ | $-1$ | $0$ | $0$ | $2$ |
| $[3]$ | $-0.4$ | $0$ | $0$ | $0$ |
| $[1]$ | $-0.4$ | $0$ | $0$ | $0$ |

# Example - Iteration 5.

$$\mathbf{x}^{[2]} = [0, 0], \quad y^{[2]} = 1$$

$$\begin{aligned} \hat{y} &= \text{sign}(0 \cdot 0 + 0 \cdot 0 - 0.4) \\ &= \text{sign}(-0.4) = -1 \end{aligned}$$

$$e = 1 - (-1) = 2 \qquad \text{UPDATE}$$

$$\begin{aligned} w_1 &\leftarrow 0 + 0.3 \cdot 2 \cdot 0 = 0 \\ w_2 &\leftarrow 0 + 0.3 \cdot 2 \cdot 0 = 0 \\ w_0 &\leftarrow -0.4 + 0.3 \cdot 2 \cdot 1 = 0.2 \end{aligned}$$

| | $w_0$ | $w_1$ | $w_2$ | $e$ |
|---|---|---|---|---|
| [1] | $-1$ | 0 | 0 | 0 |
| [2] | $-1$ | 0 | 0 | 2 |
| [3] | $-0.4$ | 0 | 0 | 0 |
| [1] | $-0.4$ | 0 | 0 | 0 |
| [2] | $-0.4$ | 0 | 0 | 2 |

# Example - Iteration 6.

$$\mathbf{x}^{[3]} = [1, 2], \quad y^{[3]} = -1$$

$$\begin{aligned} \hat{y} &= \text{sign}(0 \cdot 1 + 0 \cdot 2 + 0.2) \\ &= \text{sign}(0.2) = 1 \end{aligned}$$

$$e = -1 - 1 = -2 \qquad \text{UPDATE}$$

$$\begin{aligned} w_1 &\leftarrow 0 + 0.3 \cdot (-2) \cdot 1 = -0.6 \\ w_2 &\leftarrow 0 + 0.3 \cdot (-2) \cdot 2 = -1.2 \\ w_0 &\leftarrow 0.2 + 0.3 \cdot (-2) \cdot 1 = -0.4 \end{aligned}$$

| | $w_0$ | $w_1$ | $w_2$ | $e$ |
|---|---|---|---|---|
| [1] | $-1$ | $0$ | $0$ | $0$ |
| [2] | $-1$ | $0$ | $0$ | $2$ |
| [3] | $-0.4$ | $0$ | $0$ | $0$ |
| [1] | $-0.4$ | $0$ | $0$ | $0$ |
| [2] | $-0.4$ | $0$ | $0$ | $2$ |
| [3] | $0.2$ | $0$ | $0$ | $-2$ |

# Example - Iteration 7.

$$\mathbf{x}^{[1]} = [0, 1], \quad y^{[1]} = -1$$

$$\begin{aligned} \hat{y} &= \text{sign}((-0.6) \cdot 0 + (-1.2) \cdot 1 - 0.4) \\ &= \text{sign}(-1.6) = -1 \end{aligned}$$

$$e = -1 - (-1) = 0 \qquad \text{OK}$$

| | $w_0$ | $w_1$ | $w_2$ | $e$ |
|---|---|---|---|---|
| [1] | $-1$ | $0$ | $0$ | $0$ |
| [2] | $-1$ | $0$ | $0$ | $2$ |
| [3] | $-0.4$ | $0$ | $0$ | $0$ |
| [1] | $-0.4$ | $0$ | $0$ | $0$ |
| [2] | $-0.4$ | $0$ | $0$ | $2$ |
| [3] | $0.2$ | $0$ | $0$ | $-2$ |
| [1] | $-0.4$ | $-0.6$ | $-1.2$ | $0$ |

# Example - Iteration 8.

$$\mathbf{x}^{[2]} = [0, 0], \quad y^{[2]} = 1$$

$$\begin{aligned} \hat{y} &= \text{sign}(0 + 0 - 0.4) \\ &= \text{sign}(-0.4) = -1 \end{aligned}$$

$$e = 1 - (-1) = 2 \qquad \text{UPDATE}$$

$$\begin{aligned} w_1 &\leftarrow -0.6 + 0.3 \cdot 2 \cdot 0 = -0.6 \\ w_2 &\leftarrow -1.2 + 0.3 \cdot 2 \cdot 0 = -1.2 \\ w_0 &\leftarrow -0.4 + 0.3 \cdot 2 \cdot 1 = 0.2 \end{aligned}$$

| | $w_0$ | $w_1$ | $w_2$ | $e$ |
|---|---|---|---|---|
| [1] | $-1$ | $0$ | $0$ | $0$ |
| [2] | $-1$ | $0$ | $0$ | $2$ |
| [3] | $-0.4$ | $0$ | $0$ | $0$ |
| [1] | $-0.4$ | $0$ | $0$ | $0$ |
| [2] | $-0.4$ | $0$ | $0$ | $2$ |
| [3] | $0.2$ | $0$ | $0$ | $-2$ |
| [1] | $-0.4$ | $-0.6$ | $-1.2$ | $0$ |
| [2] | $-0.4$ | $-0.6$ | $-1.2$ | $2$ |

# Example - Iteration 9.

$$\mathbf{x}^{[3]} = [1, 2], \quad y^{[3]} = -1$$

$$\begin{aligned} \hat{y} &= \text{sign}((-0.6) \cdot 1 + (-1.2) \cdot 2 + 0.2) \\ &= \text{sign}(-2.8) = -1 \end{aligned}$$

$$e = -1 - (-1) = 0 \qquad \text{OK}$$

| | $w_0$ | $w_1$ | $w_2$ | $e$ |
|---|---|---|---|---|
| [1] | $-1$ | $0$ | $0$ | $0$ |
| [2] | $-1$ | $0$ | $0$ | $2$ |
| [3] | $-0.4$ | $0$ | $0$ | $0$ |
| [1] | $-0.4$ | $0$ | $0$ | $0$ |
| [2] | $-0.4$ | $0$ | $0$ | $2$ |
| [3] | $0.2$ | $0$ | $0$ | $-2$ |
| [1] | $-0.4$ | $-0.6$ | $-1.2$ | $0$ |
| [2] | $-0.4$ | $-0.6$ | $-1.2$ | $2$ |
| [3] | $0.2$ | $-0.6$ | $-1.2$ | $0$ |

# Example - Iteration 10.

$$\mathbf{x}^{[1]} = [0, 1], \quad y^{[1]} = -1$$

$$\begin{aligned}\hat{y} &= \text{sign}((-0.6) \cdot 0 + (-1.2) \cdot 1 + 0.2) \\ &= \text{sign}(-1.0) = -1\end{aligned}$$

$$e = -1 - (-1) = 0 \qquad \text{OK}$$

| | $w_0$ | $w_1$ | $w_2$ | $e$ |
|---|---|---|---|---|
| [1] | $-1$ | $0$ | $0$ | $0$ |
| [2] | $-1$ | $0$ | $0$ | $2$ |
| [3] | $-0.4$ | $0$ | $0$ | $0$ |
| [1] | $-0.4$ | $0$ | $0$ | $0$ |
| [2] | $-0.4$ | $0$ | $0$ | $2$ |
| [3] | $0.2$ | $0$ | $0$ | $-2$ |
| [1] | $-0.4$ | $-0.6$ | $-1.2$ | $0$ |
| [2] | $-0.4$ | $-0.6$ | $-1.2$ | $2$ |
| [3] | $0.2$ | $-0.6$ | $-1.2$ | $0$ |
| [1] | $0.2$ | $-0.6$ | $-1.2$ | $0$ |

# Example - Iteration 11.

$$\mathbf{x}^{[2]} = [0, 0], \quad y^{[2]} = 1$$

$$\begin{aligned} \hat{y} &= \text{sign}(0 + 0 + 0.2) \\ &= \text{sign}(0.2) = 1 \end{aligned}$$

$$e = 1 - 1 = 0 \qquad \text{OK}$$

| | $w_0$ | $w_1$ | $w_2$ | $e$ |
|---|---|---|---|---|
| [1] | $-1$ | $0$ | $0$ | $0$ |
| [2] | $-1$ | $0$ | $0$ | $2$ |
| [3] | $-0.4$ | $0$ | $0$ | $0$ |
| [1] | $-0.4$ | $0$ | $0$ | $0$ |
| [2] | $-0.4$ | $0$ | $0$ | $2$ |
| [3] | $0.2$ | $0$ | $0$ | $-2$ |
| [1] | $-0.4$ | $-0.6$ | $-1.2$ | $0$ |
| [2] | $-0.4$ | $-0.6$ | $-1.2$ | $2$ |
| [3] | $0.2$ | $-0.6$ | $-1.2$ | $0$ |
| [1] | $0.2$ | $-0.6$ | $-1.2$ | $0$ |
| [2] | $0.2$ | $-0.6$ | $-1.2$ | $0$ |

# Example - Iteration 12.

$$\mathbf{x}^{[3]} = [1, 2], \quad y^{[3]} = -1$$

$$\begin{aligned}\hat{y} &= \text{sign}((-0.6) \cdot 1 + (-1.2) \cdot 2 + 0.2) \\ &= \text{sign}(-2.8) = -1\end{aligned}$$

$$e = -1 - (-1) = 0 \qquad \text{OK}$$

A full epoch without updates so we can stop, the algorithm converged!

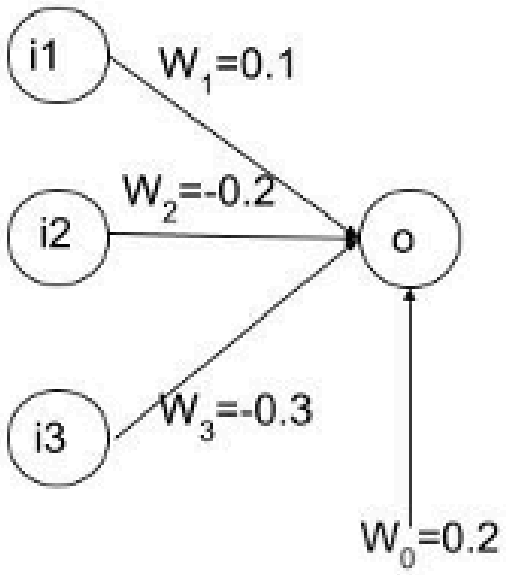
| | $w_0$ | $w_1$ | $w_2$ | $e$ |
|---|---|---|---|---|
| [1] | $-1$ | $0$ | $0$ | $0$ |
| [2] | $-1$ | $0$ | $0$ | $2$ |
| [3] | $-0.4$ | $0$ | $0$ | $0$ |
| [1] | $-0.4$ | $0$ | $0$ | $0$ |
| [2] | $-0.4$ | $0$ | $0$ | $2$ |
| [3] | $0.2$ | $0$ | $0$ | $-2$ |
| [1] | $-0.4$ | $-0.6$ | $-1.2$ | $0$ |
| [2] | $-0.4$ | $-0.6$ | $-1.2$ | $2$ |
| [3] | $0.2$ | $-0.6$ | $-1.2$ | $0$ |
| [1] | $0.2$ | $-0.6$ | $-1.2$ | $0$ |
| [2] | $0.2$ | $-0.6$ | $-1.2$ | $0$ |
| [3] | $0.2$ | $-0.6$ | $-1.2$ | $0$ |

# Exercise 1

**Exercise 2 (6 points)**

Train the following neural network until the method converges. The activation function is the function: f(x) = sign(x) and **λ**=0.2.

Reminder: at the end of each step write the updated weights. The update formula is: $W_{new} = W_{old} + \lambda \cdot (real_{label} - pred_{label}) \cdot input$

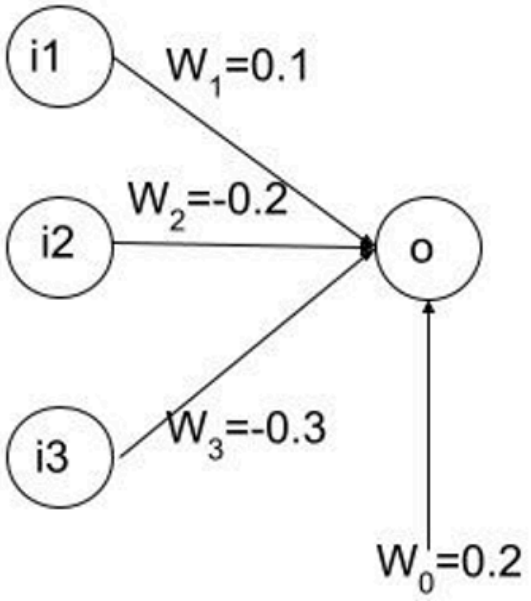



| i1 | i2 | 13 | Real label |
|---|---|---|---|
| 1 | 0 | -1 | 1 |
| 1 | 1 | 0 | -1 |
| -1 | 0 | 1 | -1 |

# Exercise 2

**Exercise 2 (6 points)**

Train the following neural network until the method converges. The activation function is the function: f(x) = sign(x+0.1) (where sign(x) is equal to +1 when x >=0, is equal to -1 otherwise) and **λ**=0.2. Reminder: the update formula is: $W_{new} = W_{old} + \lambda \cdot (real_{label} - pred_{label}) \cdot input$




| i1 | i2 | 13 | Real label |
|---|---|---|---|
| 1 | 0 | -1 | 1 |
| 0 | 1 | 0 | 1 |
| 0 | 0 | 1 | -1 |