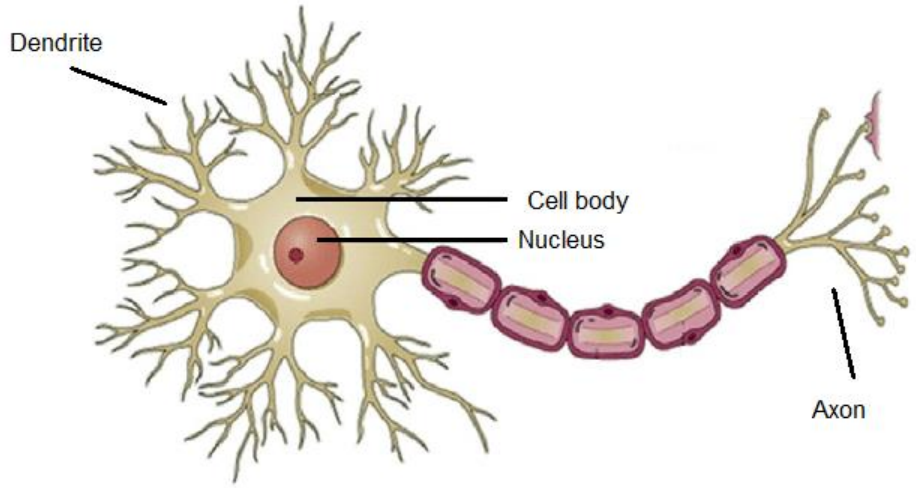
Foundations of Data Mining Neural Networks, An Introduction

Mohammad Javad Fadaeieslam

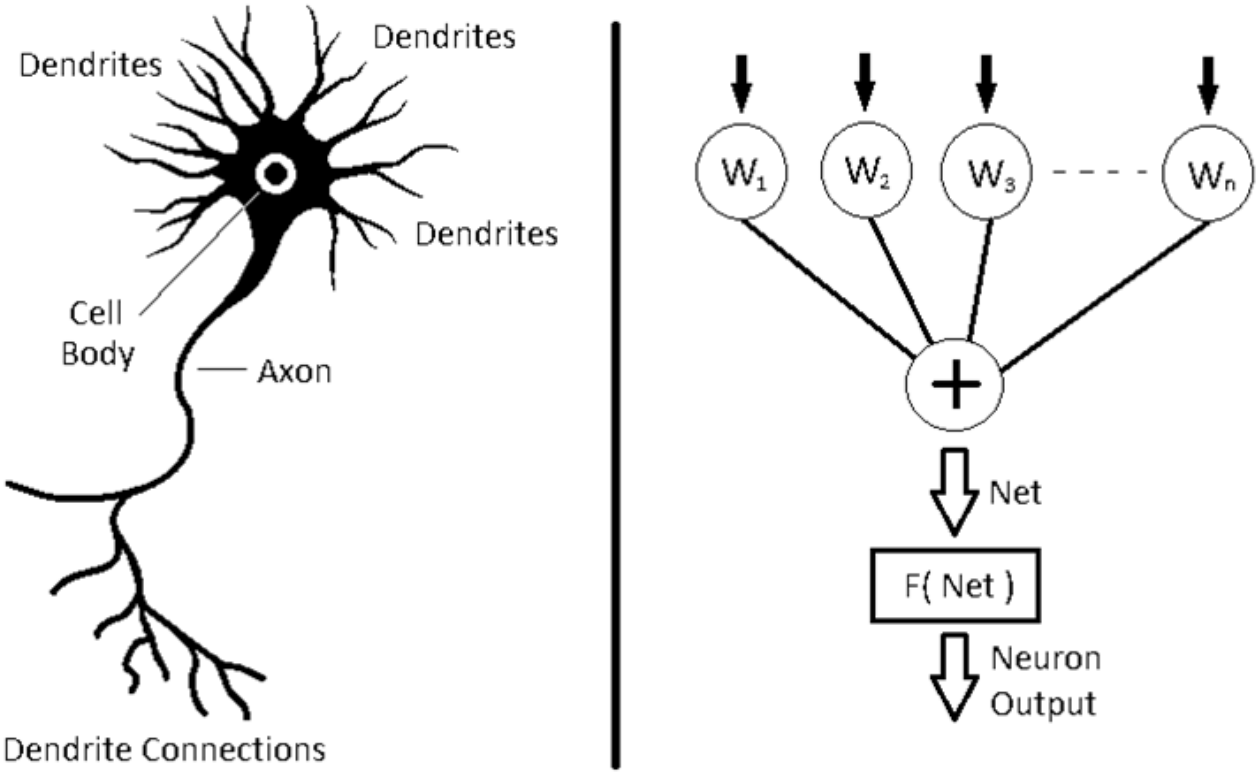
Simple Biological Neuron

- The neuron has dendrites that receive signals from other neurons.
- A cell body controls activati
- An axon carries an electric impulse to the dendrites of other neurons.

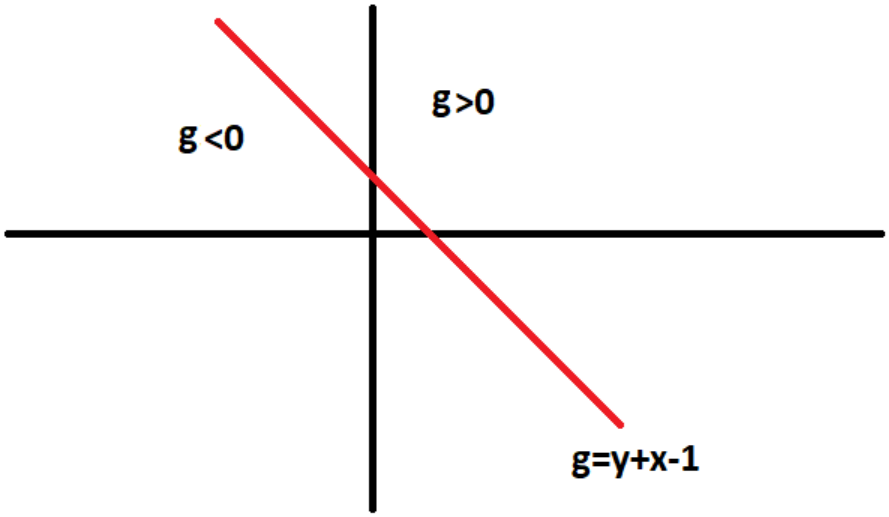


Artificial neural networks

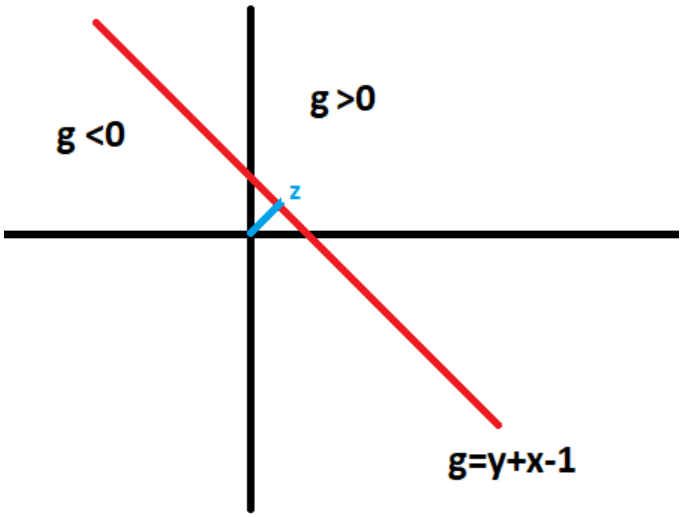
- The artificial neuron to the right has a series of weighted inputs, a summing function that groups the inputs and a **firing mechanism** (**F(Net)**), which decides whether the inputs have reached a threshold, and, if so, the neuron will fire:



معادله خط



فاصله از خط راست



معادله خطوط مختلف برای یک خط

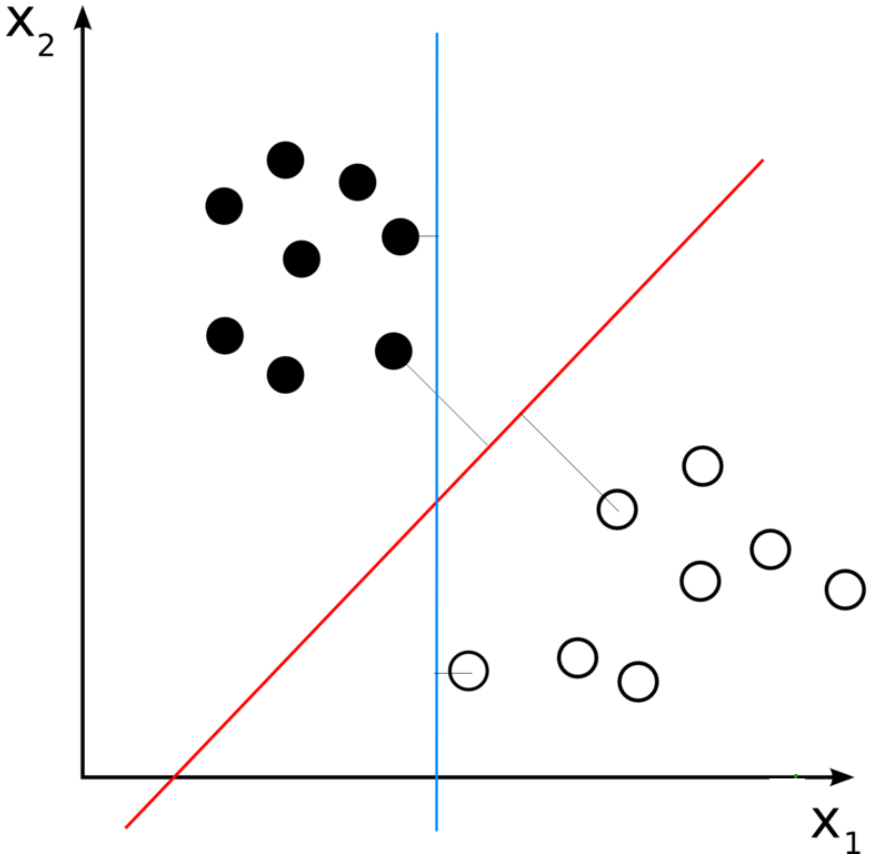
$$g = y + x - 1$$
$$g = 2y + 2x - 2$$
$$g = -y - x + 1$$

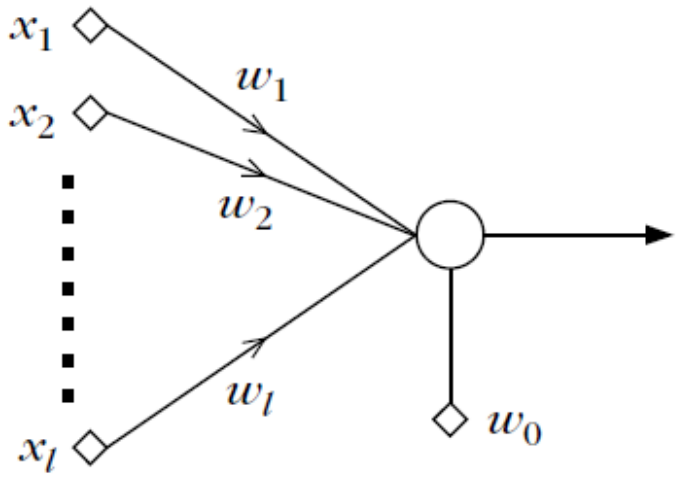
$$z = \frac{|g(\boldsymbol{x})|}{\sqrt{w_1^2 + w_2^2}}$$

$$g([0,0]) = 0 + 0 - 1 = -1 < 0$$

$$z = \frac{|-1|}{\sqrt{2}} = \frac{\sqrt{2}}{2}$$

طبقه‌بند خطی، مدل پرسپترون





(b)

The basic perceptron model

x_i ویژگی است.

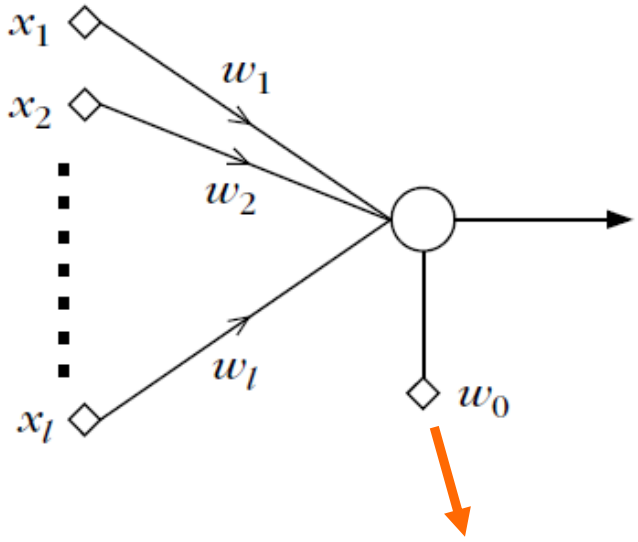
خط قرمز و آبی جداکننده داده‌های دو کلاس هستند.

فاصله نمونه‌ها از خط قرمز بیشتر است.

تابع جداکننده خطی و ابرصفحه تصمیم‌گیری

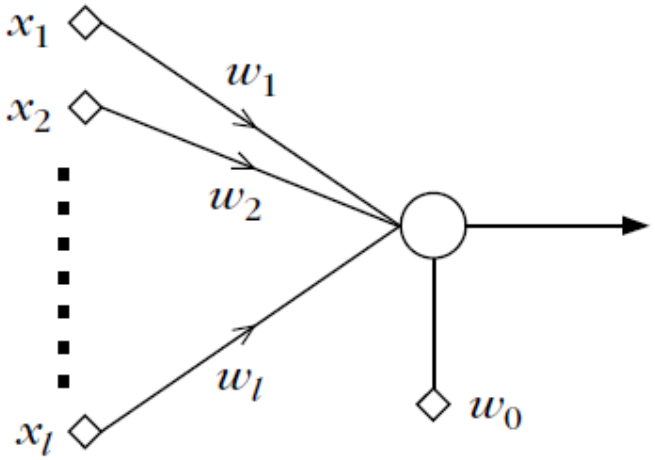
$$g = x_1 w_1 + x_2 w_2 + x_3 w_3 + \cdots + w_0$$

$$\boldsymbol{X} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_{l-1} \\ x_l \end{bmatrix}, \boldsymbol{W} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_{l-1} \\ w_l \end{bmatrix}$$



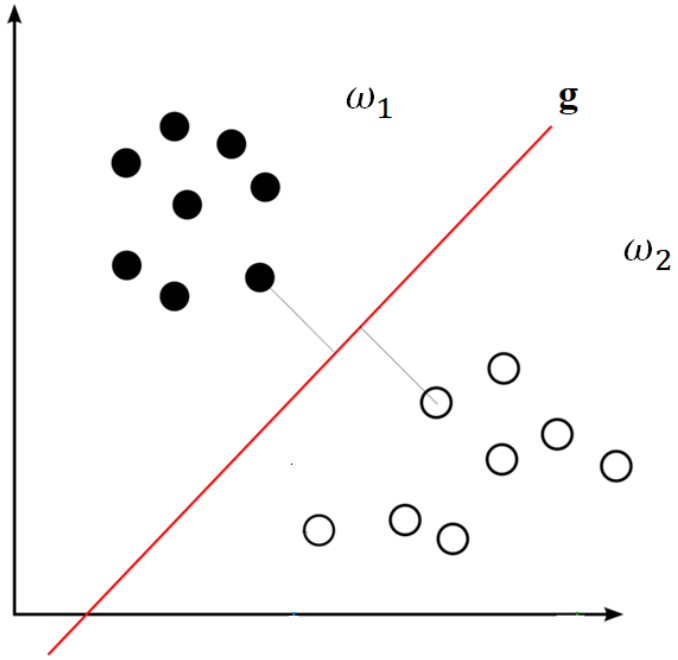
$$g = \boldsymbol{W}^T \boldsymbol{X}$$

T = transpose=ترانهاده



آموزش پرسپترون

- هدف از آموزش پرسپترون یافتن وزن‌ها (معادله خط) است به صورتی که داده‌ها درست کلاس‌بندی شوند.



$$g = x_1 w_1 + x_2 w_2 + x_3 w_3 + \cdots + w_0$$

$$\boldsymbol{W} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_{l-1} \\ w_l \end{bmatrix} = ?$$

$$g = \boldsymbol{W}^T \boldsymbol{X}$$

حالت دسته‌ای
Batch Mode

The Perceptron Algorithm

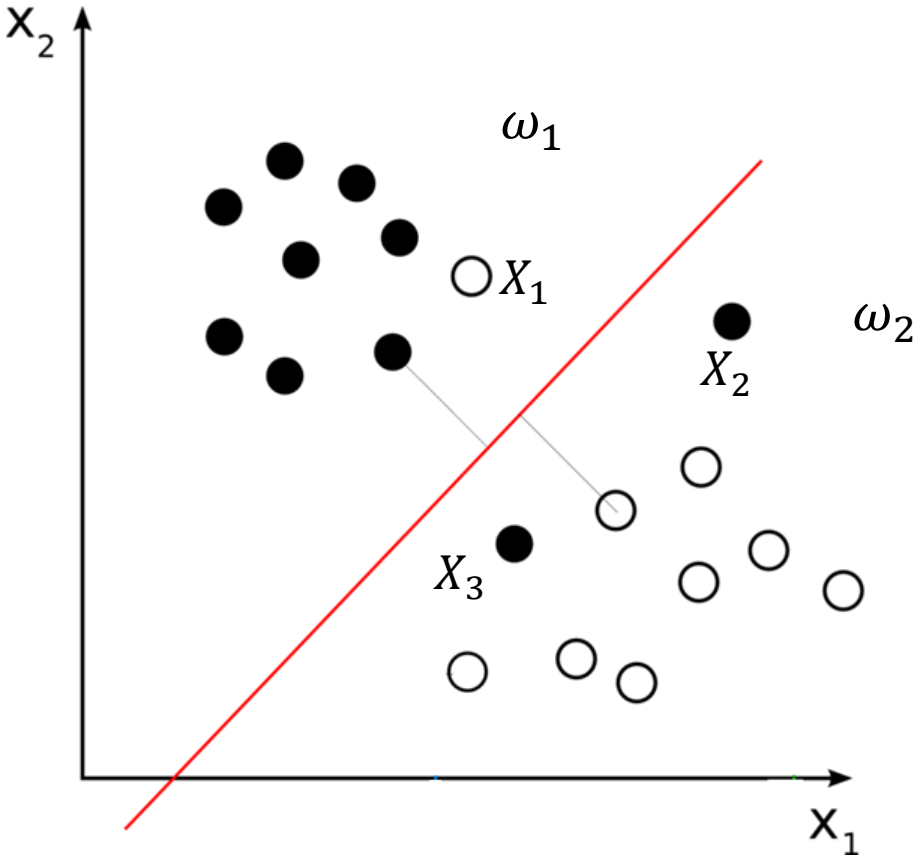
- Choose $\boldsymbol{w}(0)$ randomly
- Choose ρ_0
- $t = 0$
- Repeat
  - $Y = \emptyset$ → مجموعه نقاط با کلاس بندی نادرست
  - For $i = 1$ to N
    - If $\delta_{x_i} \boldsymbol{w}(t)^T \boldsymbol{x}_i \geq 0$ then $Y = Y \cup \{\boldsymbol{x}_i\}$

    حلقه برای یافتن مجموعه نقاط با کلاس بندی نادرست
  - End {For}
  - $\boldsymbol{w}(t+1) = \boldsymbol{w}(t) - \rho_t \sum_{x \in Y} \delta_x \boldsymbol{x}$ → به روز رسانی وزن‌ها
  - $t = t + 1$
- Until $Y = \emptyset$

$$\rho_0 = 1$$
$$\rho_t = \frac{1}{t}$$

به مرور زمان ρ کاهش می یابد.

الگوریتم آموزش پرسپترون (دسته‌ای)



$$\delta_x = -1, if\ x \in \omega_1$$
$$\delta_x = +1, if\ x \in \omega_2$$

$$Y = \{X_1, X_2, X_3\}$$

EXAMPLE

$\boldsymbol{w}(0)$: $x_1 + x_2 - 0.5 = 0 \quad \rho = 0.7$

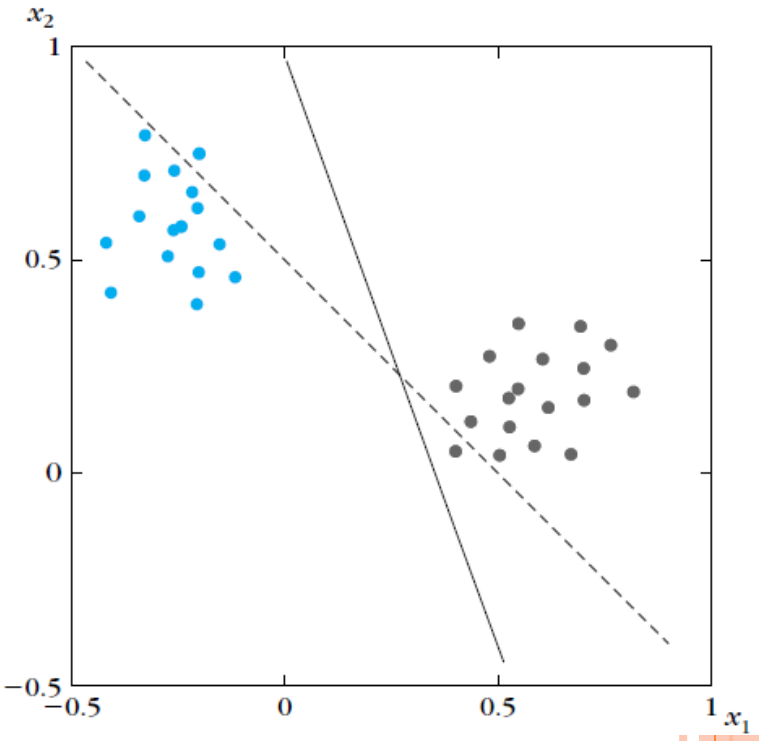
The line classifies correctly
all the vectors except $[0.4, 0.05]^T$ and $[-0.20, 0.75]^T$.
the next weight vector will be

$$\boldsymbol{w}(t+1) = \begin{bmatrix} 1 \\ 1 \\ -0.5 \end{bmatrix} - 0.7(-1) \begin{bmatrix} 0.4 \\ 0.05 \\ 1 \end{bmatrix} - 0.7(+1) \begin{bmatrix} -0.2 \\ 0.75 \\ 1 \end{bmatrix}$$

or

$$\boldsymbol{w}(t+1) = \begin{bmatrix} 1.42 \\ 0.51 \\ -0.5 \end{bmatrix}$$

The resulting new (solid) line $1.42x_1 + 0.51x_2 - 0.5 = 0$
classifies all vectors correctly, and
the algorithm is terminated.



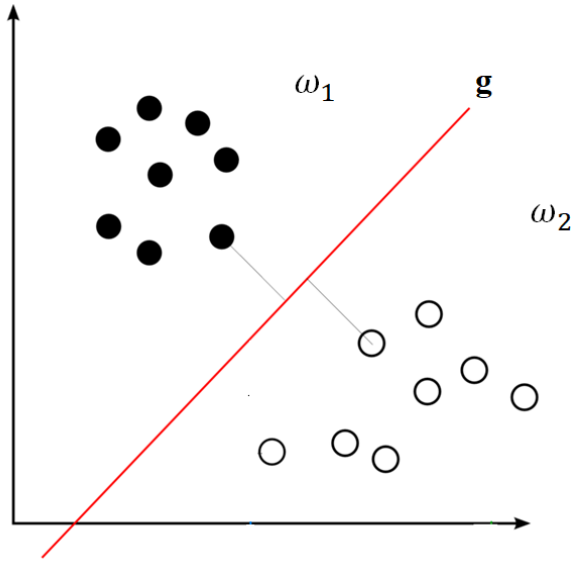
الگوریتم آموزش پرسپترون (دسته‌ای)

الگوریتم رزنبلات

$$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) + \rho \boldsymbol{x}_{(t)} \quad \text{if } \boldsymbol{x}_{(t)} \in \omega_1 \text{ and } \boldsymbol{w}^T(t)\boldsymbol{x}_{(t)} \leq 0$$

$$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) - \rho \boldsymbol{x}_{(t)} \quad \text{if } \boldsymbol{x}_{(t)} \in \omega_2 \text{ and } \boldsymbol{w}^T(t)\boldsymbol{x}_{(t)} \geq 0$$

$$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) \quad \text{otherwise}$$

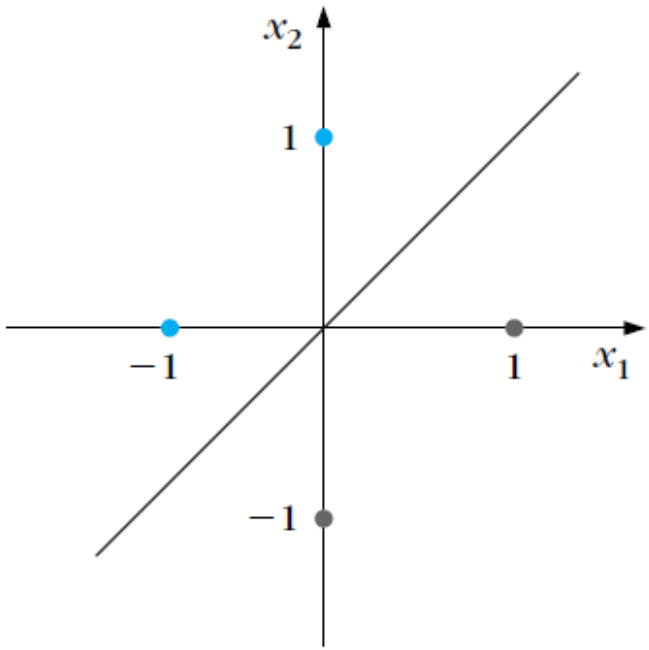


الگوریتم رزنبلات – مثال

Points $(-1,0),(0,1)$ belong to class ω_1,
and points $(0,-1),(1,0)$ belong to class ω_2.
The parameter ρ is set equal to one, and the initial weight vector is chosen as $\boldsymbol{w}(0)=[0,0,0]^T$ in the extended three-dimensional space.

Step 1. $$\boldsymbol{w}^T(0)\begin{bmatrix}-1\\0\\1\end{bmatrix}=0,\quad \boldsymbol{w}(1)=\boldsymbol{w}(0)+\begin{bmatrix}-1\\0\\1\end{bmatrix}=\begin{bmatrix}-1\\0\\1\end{bmatrix}$$

Step 2. $$\boldsymbol{w}^T(1)\begin{bmatrix}0\\1\\1\end{bmatrix}=1>0,\quad \boldsymbol{w}(2)=\boldsymbol{w}(1)$$



EXAMPLE 2 (CONT)

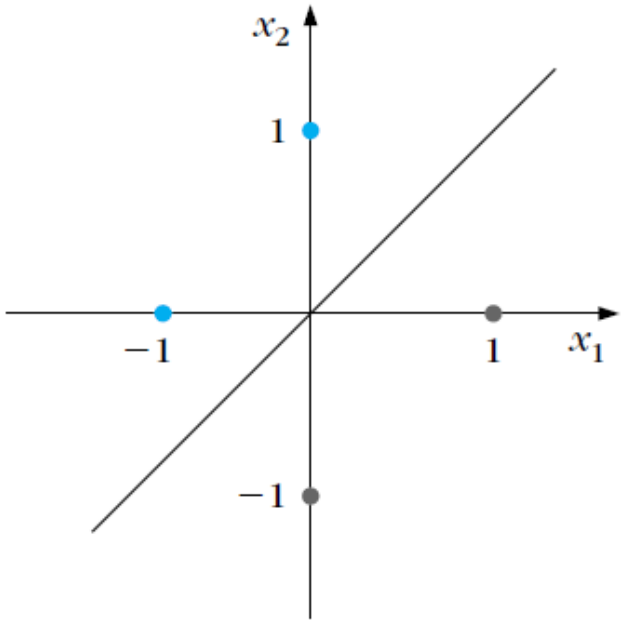
Step 3. $$\boldsymbol{w}^T(2)\begin{bmatrix}0\\-1\\1\end{bmatrix} = 1 > 0, \quad \boldsymbol{w}(3) = \boldsymbol{w}(2) - \begin{bmatrix}0\\-1\\1\end{bmatrix} = \begin{bmatrix}-1\\1\\0\end{bmatrix}$$

Step 4. $$\boldsymbol{w}^T(3)\begin{bmatrix}1\\0\\1\end{bmatrix} = -1 < 0, \quad \boldsymbol{w}(4) = \boldsymbol{w}(3)$$

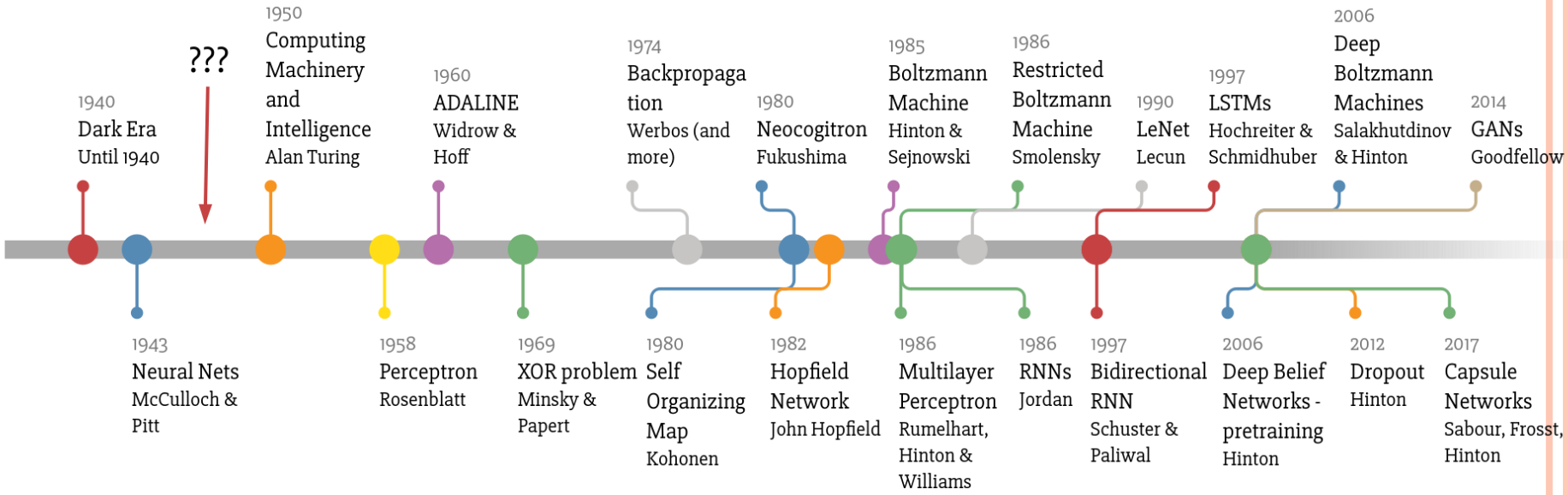
Step 5. $$\boldsymbol{w}^T(4)\begin{bmatrix}-1\\0\\1\end{bmatrix} = 1 > 0, \quad \boldsymbol{w}(5) = \boldsymbol{w}(4)$$

Step 6. $$\boldsymbol{w}^T(5)\begin{bmatrix}0\\1\\1\end{bmatrix} = 1 > 0, \quad \boldsymbol{w}(6) = \boldsymbol{w}(5)$$

Step 7. $$\boldsymbol{w}^T(6)\begin{bmatrix}0\\-1\\1\end{bmatrix} = -1 < 0, \quad \boldsymbol{w}(7) = \boldsymbol{w}(6)$$

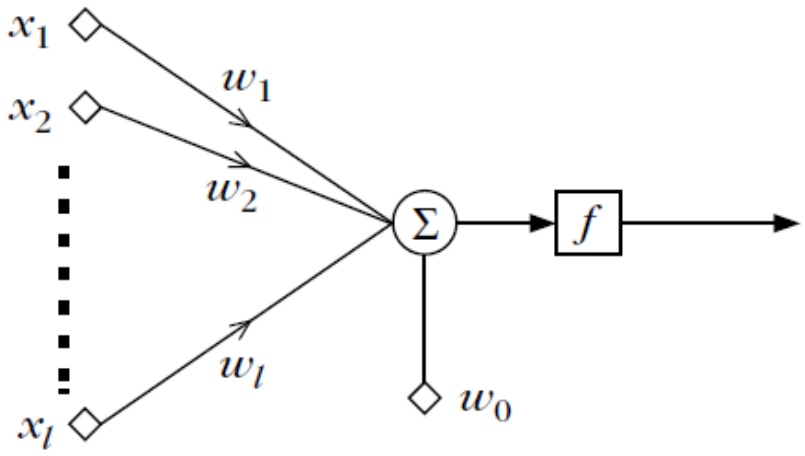


Deep Learning Timeline



Made by Favio Vázquez

THE BASIC PERCEPTRON MODEL



(a)

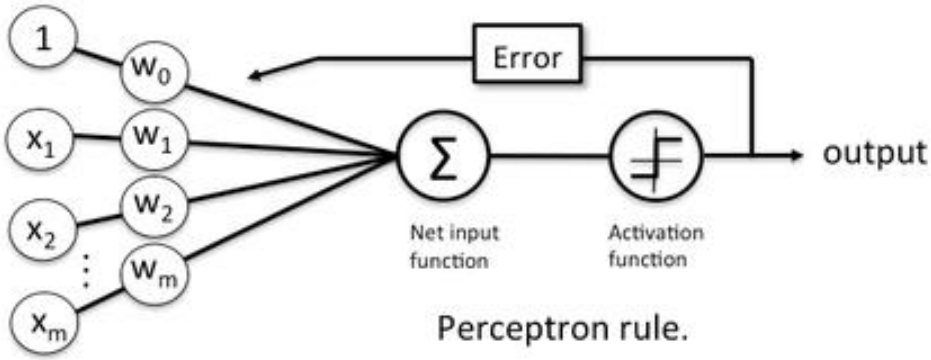
$f(\cdot)$ is the step function

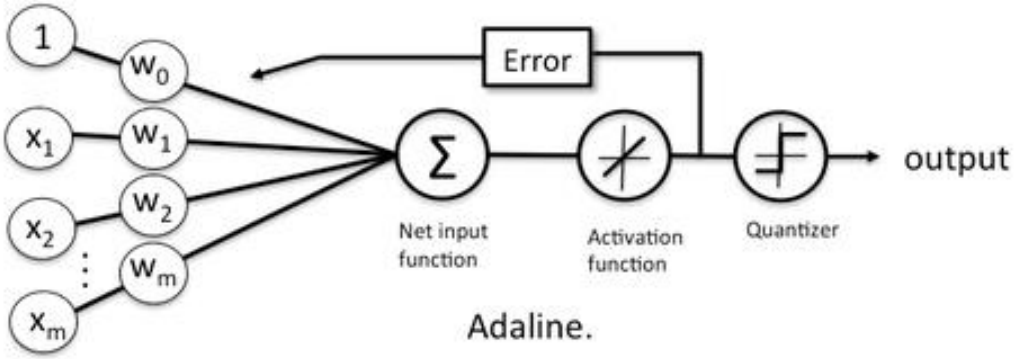
$f(x) = -1$ if $x < 0$

$f(x) = 1$ if $x > 0$

The basic perceptron model.
A linear combiner is followed by the activation function.

PERCEPTRON VS ADALINE





The Perceptron uses the class labels to learn model coefficients.

Adaline uses continuous predicted values to learn the model coefficients, which is more "powerful" since it tells us by "how much" we were right or wrong.

ADALINE (ADAPTIVE LINEAR ELEMENT)

$$\hat{\boldsymbol{w}}(k) = \hat{\boldsymbol{w}}(k-1) + \rho_k \boldsymbol{x}_k \left(y_k - \boldsymbol{x}_k^T \hat{\boldsymbol{w}}(k-1)\right)$$

The algorithm is known as Widrow–Hoff algorithm.

Adaline vs Rosenblatt

$$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) + \rho \boldsymbol{x}_{(t)} \quad \text{if } \boldsymbol{x}_{(t)} \in \omega_1 \text{ and } \boldsymbol{w}^T(t)\boldsymbol{x}_{(t)} \leq 0$$

$$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) - \rho \boldsymbol{x}_{(t)} \quad \text{if } \boldsymbol{x}_{(t)} \in \omega_2 \text{ and } \boldsymbol{w}^T(t)\boldsymbol{x}_{(t)} \geq 0$$

$$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) \quad \text{otherwise}$$

$$\hat{\boldsymbol{w}}(k) = \hat{\boldsymbol{w}}(k-1) + \rho_k \boldsymbol{x}_k \left(y_k - \boldsymbol{x}_k^T \hat{\boldsymbol{w}}(k-1)\right) \qquad y_k = \pm 1$$