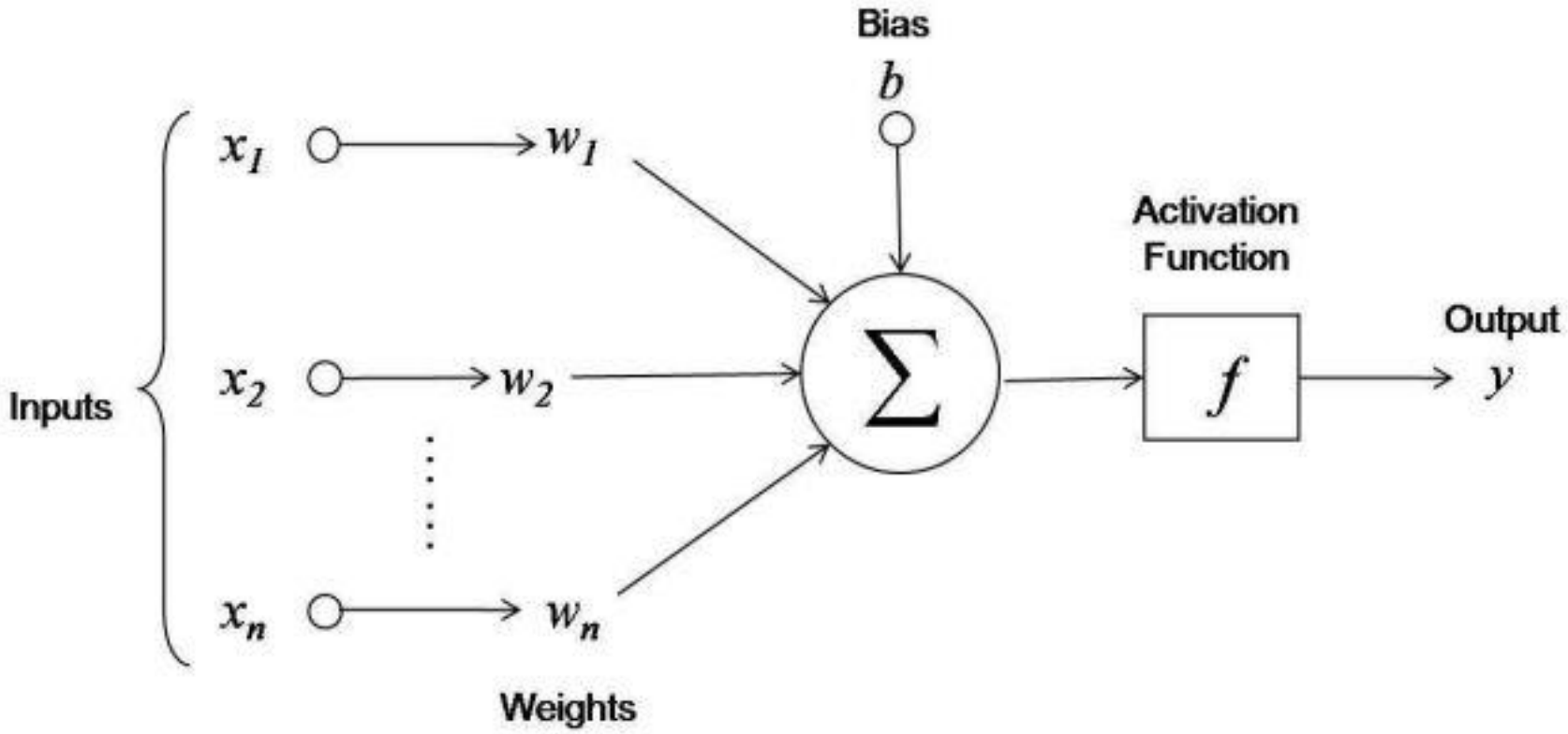
# One Layer Neural Network



- We want to move the weight vector in the direction that decrease *E*

- $w_i = w_i + \Delta w_i$ $\quad$ $w = w + \Delta w$

$$\Delta w_i = -\eta \frac{\partial E}{\partial w_i}$$

$$\Delta \vec{w} = -\eta \nabla E[\vec{w}]$$

# Gradient Descent

Gradient-Descent(*training_examples*, η)

Each training example is a pair of the form $<(x_1,\ldots x_n),t>$ where $(x_1,\ldots,x_n)$ is the vector of input values, and t is the target output value, η is the learning rate (e.g. 0.1)

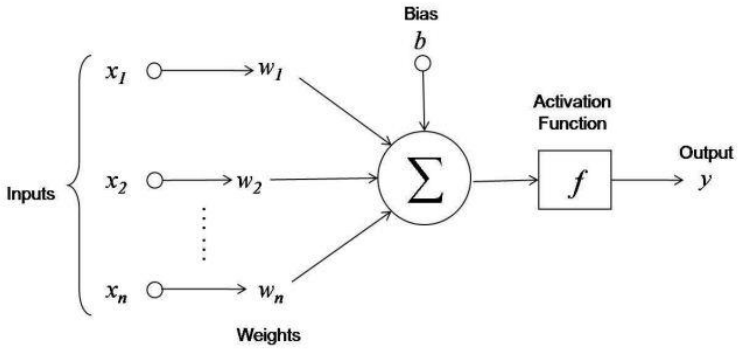
- Initialize each $w_i$ to some small random value
- Until the termination condition is meet,
- Do
  - Initialize each $\Delta w_i$ to zero
  - For each $<(x_1,\ldots x_n),t>$ in *training_examples*
  - Do
    - Input the instance $(x_1,\ldots,x_n)$ to the linear unit and compute the output o
    - For each linear unit weight $w_i$
    - Do
      - $\Delta w_i = \Delta w_i + \eta\,(t-o)\,x_i$

$$\Delta w_i = \eta \sum_{d \in D} (t_d - o_d) x_{id}$$

  - For each linear unit weight $w_i$
  - Do
    - $w_i = w_i + \Delta w_i$

# Example: Solving XOR problem with Perceptron by 2 iterations

$$\text{x\_train} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \text{targets} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$




for bias term
expand data by adding one

$$\text{x\_train} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{t} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

iter1:

$$w_0 = [0.1 \quad 0.2 \quad 0.3]$$
$$x = [1 \quad 0 \quad 1]$$
$$sum = w_0^T * x = 0.4$$
$$y = \begin{cases} 1 & \text{if } sign(sum) > 0 \\ -1 & \text{if } sign(sum) < 0 \end{cases}$$
$$y = 1$$
$$t = 1$$
$$error = 0$$
$$eta = 0.01$$
$$\Delta w = eta * error * x$$
$$\Delta w = [0.0 \quad 0.0 \quad 0.0]$$
$$w_1 = w_0 + \Delta w = [0.1 \quad 0.2 \quad 0.3]$$

iter2:

$$w_1 = [0.1 \quad 0.2 \quad 0.3]$$
$$x = [1 \quad 1 \quad 1]$$
$$sum = w_1^T * x = 0.6$$
$$y = \begin{cases} 1 & \text{if } sign(sum) > 0 \\ -1 & \text{if } sign(sum) < 0 \end{cases}$$
$$y = 1$$
$$t = 0$$
$$error = t - y$$
$$error = -1$$
$$eta = 0.01$$
$$\Delta w = eta * error * x$$
$$\Delta w = [-0.01 \quad -0.01 \quad -0.01]$$
$$w_2 = w_1 + \Delta w = [0.09 \quad 0.19 \quad 0.29]$$