$\theta = 0$ (vector)

if $y^{(i)}(\theta \cdot x^{(i)}) \le 0$ **then**

$\theta = \theta + y^{(i)}x^{(i)}$

$$y^{(i)}(\underbrace{\cancel{\theta}}_{0} + y^{(i)}x^{(i)})\cdot x^{(i)} = \|x^{(i)}\|^2 > 0$$

1. 假設起始向量為零向量>那麼接下來一定會更新到
2. 預測值乘以 label 等於 0 也算分錯的點 > 因為=0 代表在分類線上，無法得知 label

procedure PERCEPTRON($\{(x^{(i)}, y^{(i)}), i = 1, \ldots, n\}$, T)
    $\theta = 0$ (vector)
    for $t = 1, \ldots, T$ **do**
        for $i = 1, \ldots, n$ **do**
            if $y^{(i)}(\theta \cdot x^{(i)}) \le 0$ **then**
                $\theta = \theta + y^{(i)}x^{(i)}$
    return θ

Perceptron (with offset)

procedure PERCEPTRON($\{(x^{(i)}, y^{(i)}), i = 1, \ldots, n\}$, T)
    $\theta = 0$ (vector), $\theta_0 = 0$ (scalar)
    for $t = 1, \ldots, T$ **do**
        for $i = 1, \ldots, n$ **do**
            if $y^{(i)}(\theta \cdot x^{(i)} + \theta_0) \le 0$ **then**
                $\theta = \theta + y^{(i)}x^{(i)}$
                $\theta_0 = \theta_0 + y^{(i)}$
    return θ, θ_0

$$\begin{bmatrix}\theta\\ \theta_0\end{bmatrix} = \begin{bmatrix}\theta\\ \theta_0\end{bmatrix} + y^{(i)}\begin{bmatrix}x^{(i)}\\ 1\end{bmatrix}$$

$$\theta\cdot x + \theta_0 = \begin{bmatrix}\theta\\ \theta_0\end{bmatrix}\begin{bmatrix}x\\ 1\end{bmatrix}$$

我們希望 margin boundary 越大越好，所以 theta 越小越好

Perceptron algorithm: ex

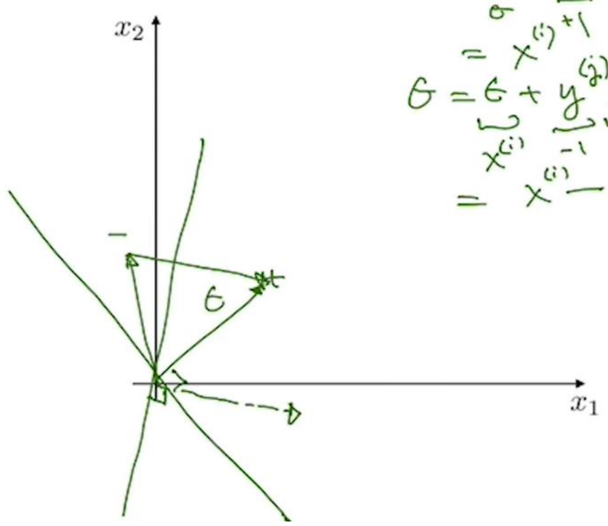
$\theta = 0$

$\theta = \underbrace{\theta}_{0} + \underbrace{y^{(i)}}_{+1} \underbrace{x^{(i)}}$

$= x^{(i)}$

$\theta = \underbrace{\theta}_{x^{(i)}} + \underbrace{y^{(j)}}_{-1} \underbrace{x^{(j)}}$

$= x^{(i)} - x^{(j)}$



The **decision boundary** is the set of points x which satisfy

$$\theta \cdot x + \theta_0 = 0.$$

The **Margin Boundary** is the set of points x which satisfy

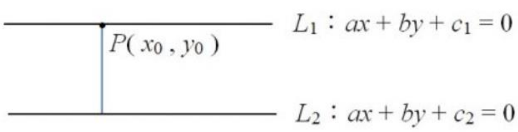
$$\theta \cdot x + \theta_0 = \pm 1.$$

So, the distance from the decision boundary to the margin boundary is $\frac{1}{||\,\theta\,||}$.

【兩平行線距離公式】

已知平面上直線 L_1：$ax + by + c_1 = 0$ 與直線 L_2：$ax + by + c_2 = 0$ 平行，則

L_1 與 L_2 的距離為 $d(L_1, L_2) = \dfrac{|c_2 - c_1|}{\sqrt{a^2 + b^2}}$



我們希望 margin boundary 越大越好，所以 theta 越小越好

- General optimization formulation of learning

  objective function = average loss + regularization

- Large margin linear classification as optimization
  - margin boundaries, hinge loss, regularization

$$J(\theta, \theta_0) = \frac{1}{n} \sum_{i=1}^{n} \text{Loss}_h\left(y^{(i)}(\theta \cdot x^{(i)} + \theta_0)\right) + \frac{\lambda}{2}\|\theta\|^2$$

用 Gradient descent 來求最佳解

$$J(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left[\text{Loss}_h(y^{(i)}\theta \cdot x^{(i)}) + \frac{\lambda}{2}\|\theta\|^2 \right]$$

sample i at random

$$\theta \leftarrow \theta - \eta \nabla_\theta \left[\text{Loss}_h(y^{(i)}\theta \cdot x^{(i)}) + \frac{\lambda}{2}\|\theta\|^2 \right]$$

$$\theta \leftarrow \theta - \eta_t \left[\begin{cases} 0, & \text{loss} = 0 \\ -y^{(i)}x^{(i)}, & \text{loss} > 0 \end{cases} + \lambda\theta \right]$$