$\theta = 0$ (vector)

if $y^{(i)}(\theta \cdot x^{(i)}) \leq 0$ 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

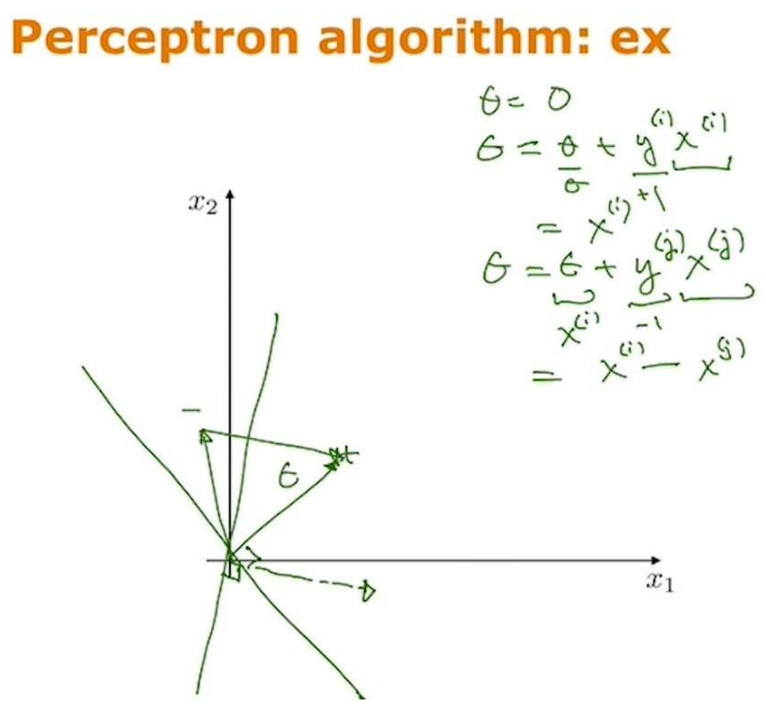
```
procedure PERCEPTRON({(x^(i), y^(i)), i = 1, ..., n}, T)
    θ = 0 (vector)
    for t = 1, ..., T do
        for i = 1, ..., n do
            if y^(i)(θ · x^(i)) ≤ 0 then
                θ = θ + y^(i) x^(i)
    return θ
```

Perceptron (with offset)

```
procedure PERCEPTRON({(x^(i), y^(i)), i = 1, ..., n}, T)
    θ = 0 (vector), θ_0 = 0 (scalar)
    for t = 1, ..., T do
        for i = 1, ..., n do
            if y^(i)(θ · x^(i) + θ_0) ≤ 0 then
                θ = θ + y^(i) x^(i)
                θ_0 = θ_0 + y^(i)
    return θ, θ_0
```

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

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



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

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

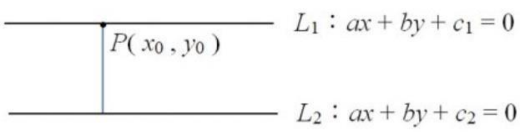
So, the distance from the decision boundary to the margin boundary is $\frac{1}{||\theta||}$.

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

【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 $d(L_1, L_2) = \frac{|c_2 - c_1|}{\sqrt{a^2 + b^2}}$



- General optimization formulation of learning

  $\underline{\text{objective function}} = \underline{\text{average loss}} + \underline{\text{regularization}}$

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

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

用 Gradient descent 來求最佳解

Large margin as optimization

CSAIL

- Hinge loss

$$\text{Loss}_h\big(\underbrace{\overbrace{y^{(i)}(\theta\cdot x^{(i)}+\theta_0)}^{\text{agreement}}}_{z}\big) = \begin{cases} 0 & \text{if } z \ge 1 \\ 1-z & \text{if } z < 1 \end{cases}$$

- Regularization: towards max margin

$$\max \frac{1}{\|\theta\|} \qquad \min \frac{1}{2}\|\theta\|^2$$

- The objective

reg. param $\lambda > 0$

$$J(\theta,\theta_0) = \underbrace{\frac{1}{n}\sum_{i=1}^{n} \text{Loss}_h\big(y^{(i)}(\theta\cdot x^{(i)}+\theta_0)\big)}_{\text{ave loss}} + \underbrace{\frac{\lambda}{2}\|\theta\|^2}_{\text{reg}}$$

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

sample i at random

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

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