算法流程

输入 $\{(x_i, y_i)\}$ $i=1\sim N$

① 随机取 (w, b)

② 随机取 x_i

ⓐ 若 $w^T x_i + b \geq 0$ 且 $y=-1$ 则 $w = w - x$, $b = b-1$

ⓑ 若 $w^T x_i + b < 0$ 且 $y=1$ 则 $w = w + x$, $b = b+1$

③ 回到 ② 直到所有 x_i ⓐ与ⓑ不成立时退出

定义一个增广向量 $\vec{X}$

ⓐ 若 $y=+1$ 则 $\vec{X} = \begin{bmatrix} x \\ 1 \end{bmatrix}$

ⓑ 若 $y=-1$ 则 $\vec{X} = \begin{bmatrix} -x \\ -1 \end{bmatrix}$

定义一个增广 W 如下:

$$W = \begin{bmatrix} w \\ b \end{bmatrix}$$

输入 $\vec{X_i}$

① 随机取 W

② 随机取 $\vec{X_i}$

若 $W^T \vec{X_i} < 0$ 则 $W = W + \vec{X}$

③ 回到 ② 直到所有 x_i ⓐ与ⓑ不成立时退出

感知器算法收敛定理

输入 $\{\vec{X_i}\}_{i=1\sim N}$ 若线性可分，即 $\exists W_{opt}$ 使:

$$W^T_{opt} \vec{X_i} > 0 \quad (i=1\sim N)$$

则利用上述感知器算法经过有限步后得到一个 W 使：

$$W^T \vec{X_i} > 0 \quad (i = 1 \sim N)$$

证明：W 不一定等于 W_{opt}

1. 不失一般性，设 $\|W_{opt}\| = 1$ （因为 W_{opt} 与 aW_{opt} 是同一平面 $a>0$）
2. 假设第 k 步的 W 是 $W(k)$

且有一个 $\vec{X_i}$ 使 $W(k)^T \vec{X_i} < 0$

则据感知器算法 $W(k+1) = W(k) + \vec{X_i}$

$$\|W(k+1) - aW_{opt}\|^2 = \|W(k) + \vec{X_i} - aW_{opt}\|^2$$

$$= \|(W(k) - aW_{opt}) + \vec{X_i}\|^2$$

$$= \|(W(k) - aW_{opt}\|^2 + \|\vec{X_i}\|^2 + 2\underbrace{W(k)^T \vec{X_i}}_{<0} - 2\underbrace{aW_{opt}^T X_i}_{>0}$$

一定可以取很大的 a 使

$$\|W(k+1) - aW_{opt}\|^2 < \|W(k) - aW_{opt}\|^2$$

定义：$\beta = \max\limits_{i=1\sim N} \{\|\vec{X_i}\|\}$

$$\gamma = \min\limits_{i=1\sim N} \{W_{opt}^T X_i\}$$

取 $a = \frac{\beta^2 + 1}{2\gamma}$ 则

$$\|W(k+1) - aW_{opt}\|^2 \le \|W(k) - aW_{opt}\|^2 - 1$$

取 $D=\|W(0)-aW_{opt}\|$ 则至多经过 D^2 步

W 将会收敛至 aW_{opt}