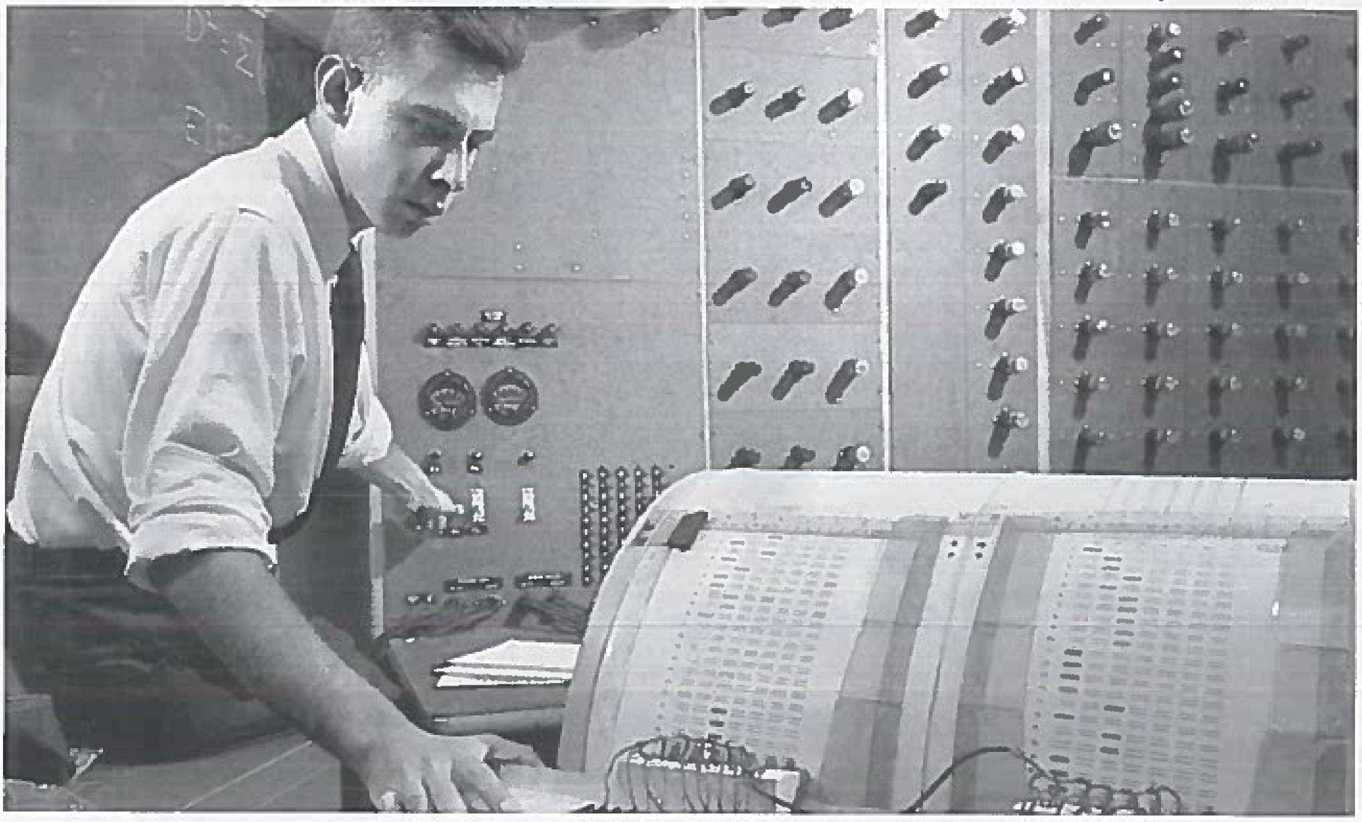
Frank Rosenblatt: Creator of the Perceptron

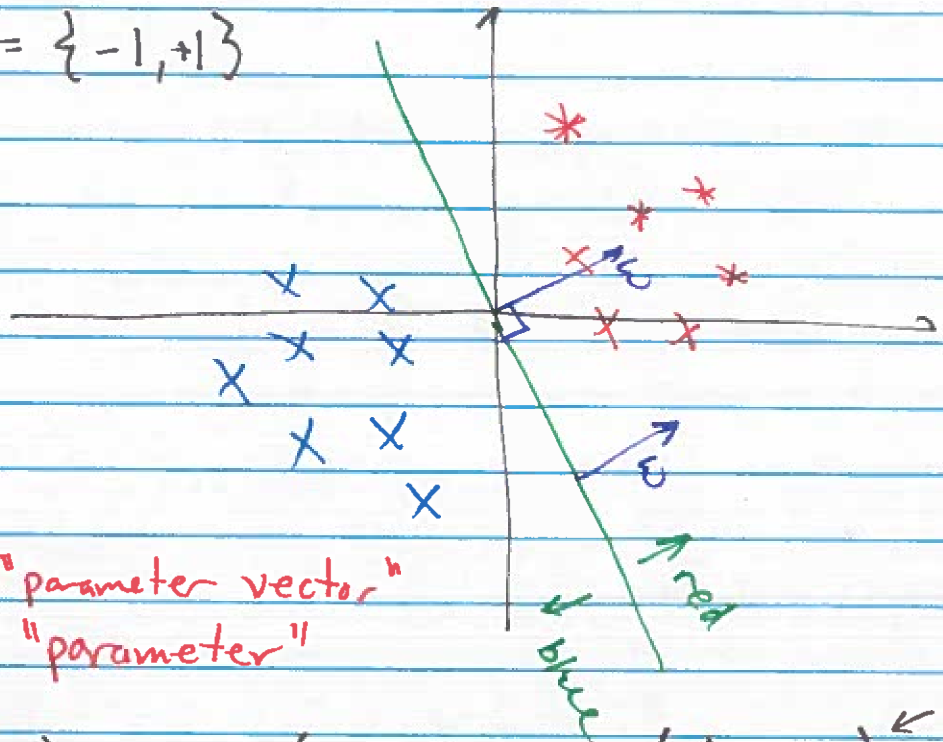


Cornell BA '50, PhD '56, and Professor

"Father of Deep Learning"

Lecture 6 - The Perceptron

$\mathcal{Y} = \{-1, +1\}$



"parameter vector"

$w \Leftarrow$ "parameter"

$$h_w(x) = \text{sign}(w^T x) = \text{sign}(\langle w, x \rangle)$$

(or) $$\text{sign}(w^T x + b) = \text{sign}\left(\left\langle \begin{bmatrix} w \\ b \end{bmatrix}, \begin{bmatrix} x \\ 1 \end{bmatrix} \right\rangle\right)$$

if $x \in \mathbb{R}^d$, then $w \in \mathbb{R}^d$

$$\begin{bmatrix} x \\ 1 \end{bmatrix} \in \mathbb{R}^{d+1} \qquad \begin{bmatrix} w \\ b \end{bmatrix} \in \mathbb{R}^{d+1}$$

$\text{sign}: \mathbb{R} \to \mathbb{R}$. $\quad \text{sign}(x) = \begin{cases} +1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \end{cases}$

$$\text{sign}(x) = \frac{x}{|x|}$$

Hypothesis class for Perceptron:

$\cancel{\{h \in \mathbb{R}^n \to \mathbb{R}^m}$ |

$$\mathcal{H} = \{h : \mathbb{R}^d \to \{+1, -1\} \mid h(x) = \text{sign}(w^T x),\ \underset{w \neq 0}{w \in \mathbb{R}^d}\}$$

Perceptron Algorithm:

$w \leftarrow 0$
loop:
  for $(x, y) \in \mathcal{D}$:
    ~~let $\hat{y} = \text{sign}(w^T x)$~~
    ~~if $\hat{y} \neq y$~~: if $y(w^T x) \leq 0$:
      $w \leftarrow w + yx$
    end if
  end for
until everything is correctly classified

$$w' = w + yx$$

$$y((w')^T x) = y((w + yx)^T x) = y w^T x + x^T x > y(w^T x)$$

Suppose there is just 1 data point? $(x,$

How many times can we misclassify starting at $w = 0$?

$w = 0.$

$y(x^T w) = 0 \leq 0$

$\Downarrow$

$w = x$

$y(w^T x) = \cancel{y(x^T x)}\ y(x^T x) \geq 0$ ✓

What if we start at some $w = w_0$?

$w = w_0$

$y(x^T w) \leq 0$ wrong

$\Downarrow$

$w = w_0 + x$

$y(x^T w) \leq 0$ wrong

$\Downarrow$

$\cancel{w = w_0 + 2x}$ $w = w_0 + 2x$

$\Downarrow$

$\cdots$

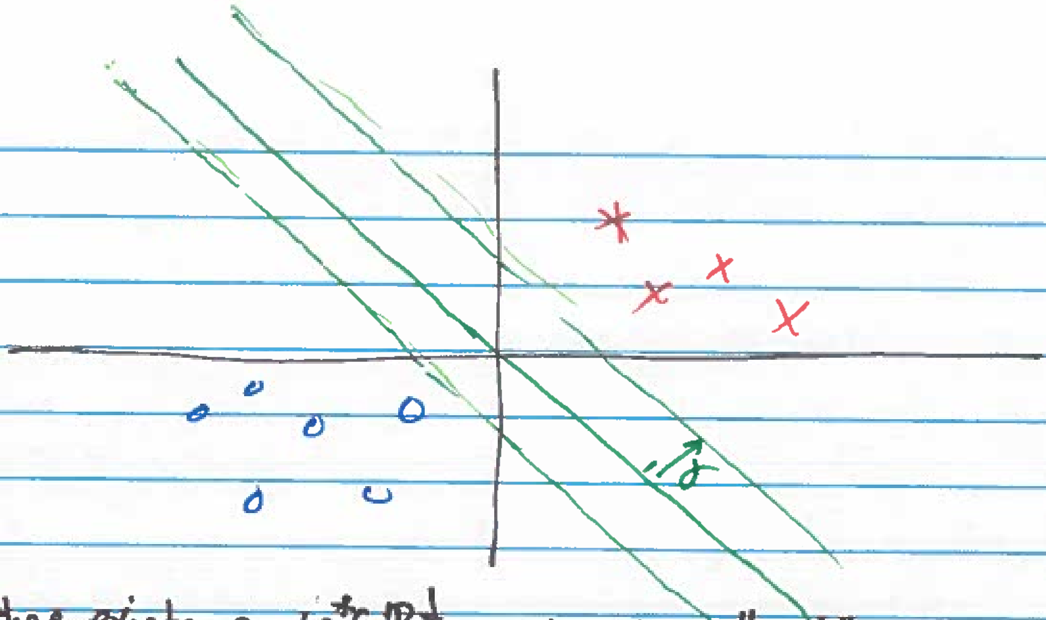
$\Downarrow$

$w = w_0 + kx$

$y(w^T x) = y(w_0^T x + k x^T x) \overset{?}{\leq} 0$

$\cancel{= y w_0^T x + k y x^T x}$

$\cancel{k \leq \frac{-y w_0^T x}{x^T x}}$



there exists a $w^* \in \mathbb{R}^d$ such that $\|w^*\| = 1$

if $(x, y) \in \mathcal{D}$, then $x^T w^* y \geq \gamma$

if $y = +1$, then $x^T w^* \geq \gamma$; if $y = -1$ $x^T w^* \leq -\gamma$

also assume $\|x\| \leq 1$.

let w_t denote the value of w after t updates.

$$\langle w_{t+1}, w^* \rangle = \langle w_t + yx, w^* \rangle = \langle w_t, w^* \rangle + yx^T w^*$$

$$\geq \langle w_t, w^* \rangle + \gamma$$

$$\langle w_t, w^* \rangle \geq \underbrace{\langle w_0, w^* \rangle}_{=0} + \gamma t = \gamma t$$

$\langle w_{t+1}, w_{t+1} \rangle = \langle w_{t+1}^{*} yx,\ w$

$$\langle w_{t+1}, w_{t+1} \rangle = \langle w_t + yx, w_t + yx \rangle$$

$$= \langle w_t, w_t \rangle + \underbrace{2yx^T w_t}_{\leq 0} + \langle yx, yx \rangle$$

$$\leq \langle w_t, w_t \rangle + \langle x, x \rangle = \|x\|^2 + \|w_t\|^2$$

$$\leq \langle w_t, w_t \rangle + 1$$

$$\langle w_t, w_t \rangle \leq \langle w_0, w_0 \rangle + t = t$$

$$\gamma t \leq \langle w_t, w^* \rangle = \|w_t\| \cdot \underset{\leq 1}{\|w^*\|} \underset{\leq 1}{\cos\theta}$$

$$\leq \|w_t\| = \sqrt{\langle w_t, w_t \rangle} \quad 6$$

$$\leq \sqrt{t}$$

$$\gamma t \leq \sqrt{t} \Rightarrow \gamma^2 t^2 \leq t \Rightarrow \gamma^2 t \leq 1$$

$$\Downarrow$$

$$t \leq \frac{1}{\gamma^2}$$

$y=1$

$$y(\omega^T x) = y(\omega_0^T x + k x^T x) \overset{?}{\leq} 0$$

$$\cancel{\omega_0^T x \leq} \quad k x^T x \leq -\omega_0^T x$$

if $k \leq \frac{-\omega_0^T x}{x^T x}$ then we're wrong.

#of times we're wrong.

at most $\max\left(\left\lfloor \frac{-\omega_0^T x}{x^T x} \right\rfloor + 1, 0\right)$

Just $\{(x,+1)\} = D$