Theorem Assume $\exists\, w^*$ s.t. for any (x,y) in the sequence

① $\text{sign}(\langle x, w^* \rangle) = y$

② $\left| \frac{\langle x, w^* \rangle}{\|w^*\|_2} \right| \geq \gamma$

③ $\|x\|_2 \leq R$

then the number of mistakes M perceptron makes obeys

$$M \leq \frac{R^2}{\gamma^2}$$

Proof: Claim A: $\|W_{M+1}\|_2 \geq M\gamma$

$$\|w^*\|_2 \|W_{M+1}\|_2 \geq \langle W^*, W_{M+1} \rangle \stackrel{\text{update rule}}{=} \langle w^*, \sum_{t=1}^{M} y_t x_t \rangle = \sum_{t=1}^{M} y_t \langle w^*, x_t \rangle$$

$$\geq M \cdot \gamma \cdot \|w^*\|_2$$

Cauchy-Shwartz Inequality

$\langle a, b \rangle \leq \|a\|_2 \|b\|_2$

by the margin assumption ① and ②

Cancelling $\|W^*\|_2$ on both sides we get $\|W_{M+1}\|_2 \geq M\gamma$

Claim B: $\|W_{M+1}\|_2 \leq \sqrt{M} \cdot R$

$$\|W_{M+1}\|^2 = \|W_M + y_M x_M\|^2 = \|W_M\|^2 + \|y_M x_M\|^2 + 2 y_M \langle W_M, x_M \rangle$$

$$= \|w_{M-1}\|^2 + \|y_M x_M\|^2 + \|y_{M-1} x_{M-1}\|^2 + 2y_M \langle w_M, x_M \rangle + 2y_{M-1} \langle w_{M-1}, x_{M-1} \rangle$$

$\cdots$

$$= \underbrace{\|w_1\|^2}_{= 0} + \underbrace{\sum_{t=1}^{M} \|y_t x_t\|^2}_{= \sum_{t=1}^{M} \|x_t\|^2 \le MR^2} + 2\sum_{t=1}^{M} \underbrace{y_t \langle w_t, x_t \rangle}_{\le 0}$$

Why?
h_{w_t} made a mistake on x_t

$$\le M R^2$$

take square root on both sides

$$\|w_{M+1}\| \le \sqrt{M} \cdot R$$

Combine Claim A and B, we have

$$M\gamma \le \|w_{M+1}\| \le \sqrt{M} R$$

$$\Rightarrow \sqrt{M}\gamma \le R \Rightarrow M \le \frac{R^2}{\gamma^2} \qquad \square$$
