有關 perceptron 收斂問題的推導

--為何 Wf 和 Wt 的內積是根號 T + 常數?

井民全, Jing, mqjing@gmail.com

Google doc: This Document

有關 perceptron 收斂問題的推導, 很多人早就知道了. 我只是有興趣把它推導一次, 記錄一下過程.

Notation

- $x_{n(t)}$:這一次做錯的點 **n**
- y_n: 點**n** 的正確答案
- w_t: 目前這次的 weight 值
- w_{t+1}: 下一次的 weight 值
- w_f: 目標 weight 值, 所形成的線能完美分割 所有 x 點到正確的 y

觀念

概念是指經過 多次次學習之後, 學習後的向量 w_t^T 與 我們的學習目標向量 w_f^T 的相似程度應該要越來越高. 衡量兩個向量的相似程度, 可以用內積 $w_f^T w_t$ 表示. 兩個向量的內積數值越大 , 表示越相像 (內積最大為 1).

令 w_f 為目標向量, 可以形成一條完美的線使得每一個 x_n 代入都能正確的切開. 即 $y_n = sign(w_f^T x_n)$.

- $w_f^T x_n$ 代表點 x_n 距離線有多遠.
- 最接近線的點, 其距離可以用 $\min_n \quad w_f^T x_n$ 表示
- 因此所有點 $x_{n(t)}$ 與線的距離, 一定滿足下式 $y_{n(t)} w_f^T x_{n(t)} \geq \min_n \quad y_n w_f^T x_n$

下一次的更新 w_{t+1}的策略是 原來的 w_t 和 $y_n x_{n(t)}$ 相加, 以縮小錯誤. 因此下一個 weight

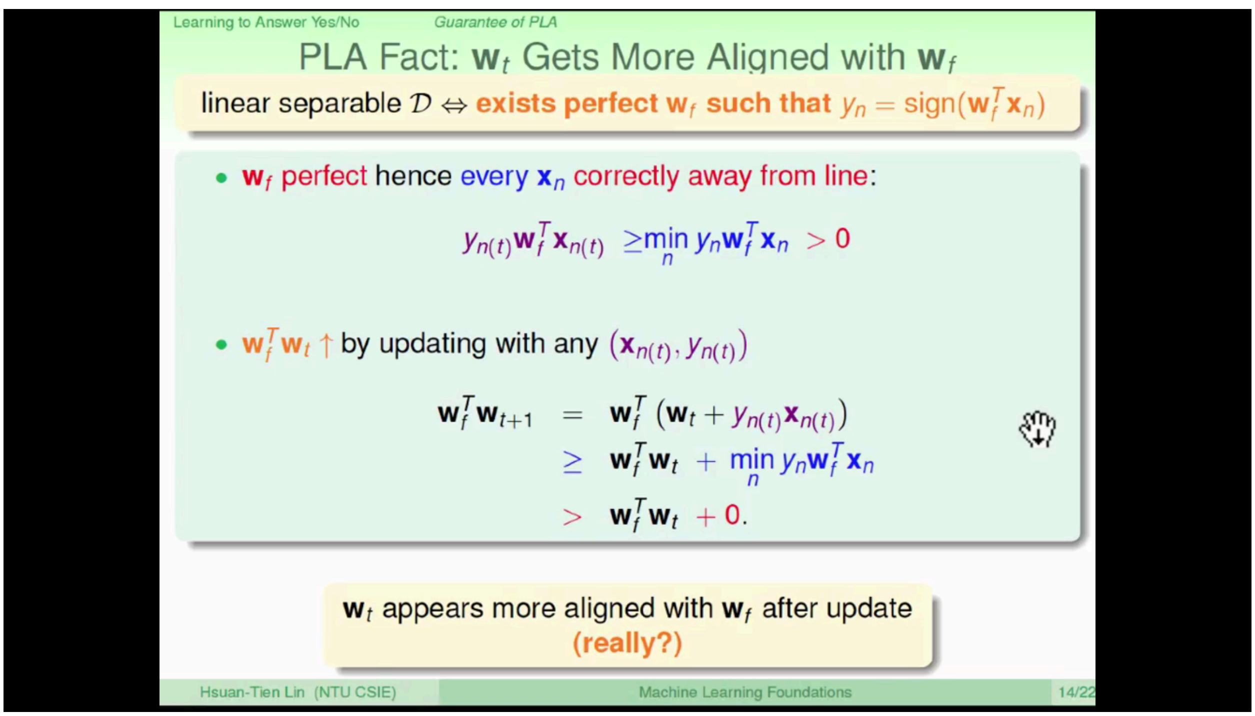
$w_{t+1} = w_t + y_n x_{n(t)}$ [ref]

每一次更新 w_{t+1}, 就應該與 w_f 更接近. 因此

$w_f^T w_{t+1}$ = $w_f^T (w_f + y_{n(t)} x_{n(t)})$

$$\geq w_t^T w_t + \min_n \quad y_n w_f^T x_n$$

如 Page 14, 顯示的結果



證明每次更新的 weight 值有 low bound 且最大值為 1.

使用遞迴算出 t 次更新 W 後, 預測 w_t^T 與 目標 w_f^T 的相似程度, 即計算內積 $w_f^T w_t$ 做法:

$$w_f^T w_{t+1} \geq w_f^T w_t + \min_n \quad y_n w_f^T x_n$$

w_f 是目標 weight.
=>
t = 0
$w_0 = 0$

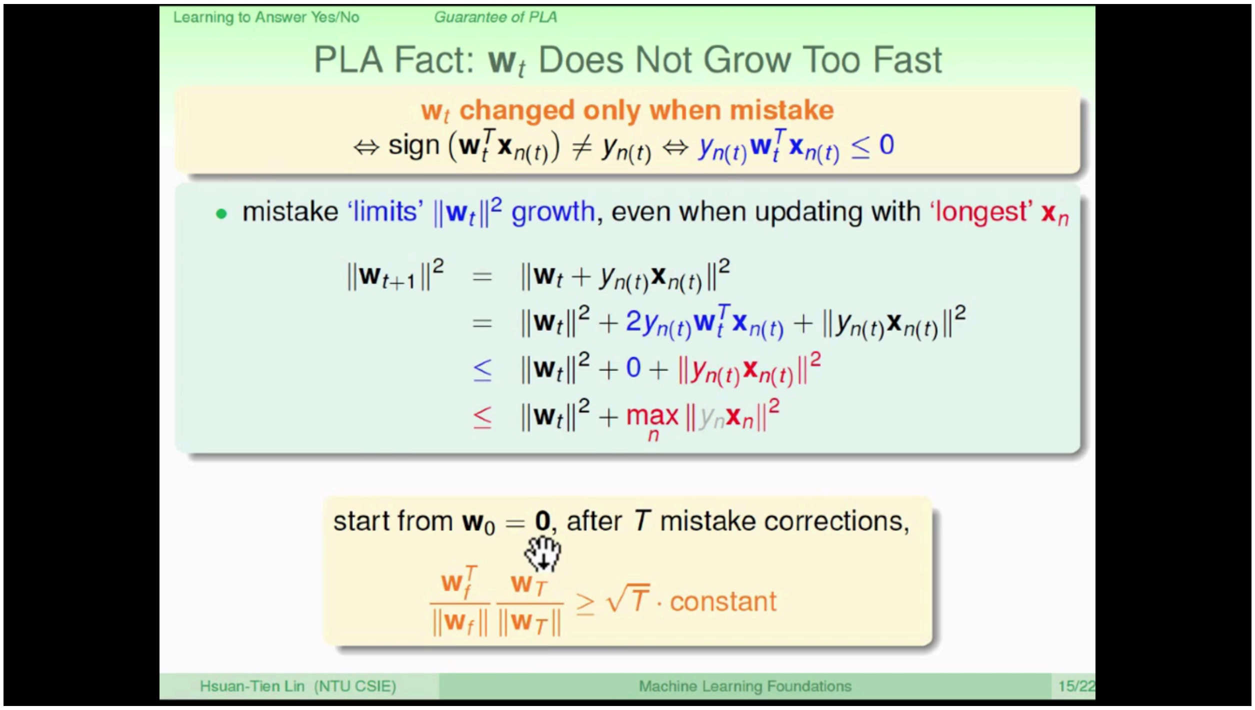
t = 1
$$w_f^T w_1 \geq w_f^T w_0 + \min_n \quad y_n w_f^T x_n$$

t = 2
$$w_f^T w_2 \geq w_f^T w_1 + \min_n \quad y_n w_f^T x_n$$
$$w_f^T w_2 \geq \min_n \quad y_n w_f^T x_n + \min_n \quad y_n w_f^T x_n$$
$$w_f^T w_2 \geq 2 \cdot \min_n \quad y_n w_f^T x_n$$

t =3
$$w_f^T w_3 \geq 3 \cdot \min_n \quad y_n w_f^T x_n$$

第 t 次
$w_f^T w_t \geq t \cdot \min_n \quad y_n w_f^T x_n$ --------- (1)

每次更新的 weight 有 upper bound.使得兩個單位向量的內積值大於根號 T倍的常數。因此可知，每次更新新都會讓 weight 更接近 w_f.



使用遞迴算出 t 次更新 W 後, 預測線 w_t 的長度變更, upper bound
做法:

$$\|w_{t+1}\|^2 \leq \|w_t\|^2 + \max_n \|y_n x_n\|^2$$

t = 0,

$$\|w_0\|^2 = 0$$

t = 1,

$$\|w_1\|^2 \leq \|w_0\|^2 + \max_n \|y_n x_n\|^2$$

$$\|w_1\|^2 \leq 0 + \max_n \|y_n x_n\|^2$$

t = 2,

$$\|w_2\|^2 \leq \|w_1\|^2 + \max_n \|y_n x_n\|^2$$

$$\|w_2\|^2 \leq \max_n \|y_n x_n\|^2 + \max_n \|y_n x_n\|^2$$

$$\|w_2\|^2 \leq 2 \cdot \max_n \|y_n x_n\|^2$$

第 t 次

$\|w_t\|^2 \leq t \cdot \max_n \quad \|y_n x_n\|^2$ -------- (2)

計算內積 $\dfrac{w_f^T}{\|w_f\|} \cdot \dfrac{w_t}{\|w_t\|}$

令

$R^2 = \max_n \|x_n\|^2$

$\rho = \min_n \quad y_n \quad \dfrac{w_f^T}{\|w_f\|} x_n$

做法如下

$\dfrac{w_f^T}{\|w_f\|} \cdot \dfrac{w_t}{\|w_t\|}$ 代入 (1) $w_f^T w_t \geq t \cdot \min_n \quad y_n w_f^T x_n$

$$\frac{w_f^T \cdot w_t}{\|w_f\| \, \|w_t\|} \geq \frac{t \cdot \min_n \quad y_n w_f^T x_n}{\|w_f\| \cdot \|w_t\|}$$

$$\geq \frac{t \cdot \rho \cdot \|w_f\|}{\|w_f\| \, \|w_t\|}$$

$$\geq \frac{t \cdot \rho}{\|w_t\|}$$

由 (2) 得知, $\|w_t\|^2 \leq t \cdot \max_n \quad \|y_n x_n\|^2$

$$\frac{1}{\|w_t\|} \geq \sqrt{\frac{1}{t \cdot \max_n \quad \|y_n x_n\|^2}}$$

, 因為 y_n 不是 (+1) 就是 (-1), 所以 $\|y_n\|$=1.

$$\geq \sqrt{\frac{1}{t \cdot R^2}}$$ --------------------(3)

$$\frac{w_f^T \cdot w_t}{\|w_f\| \, \|w_t\|} \geq \frac{t \cdot \rho}{\|w_t\|}$$

代入 (3)

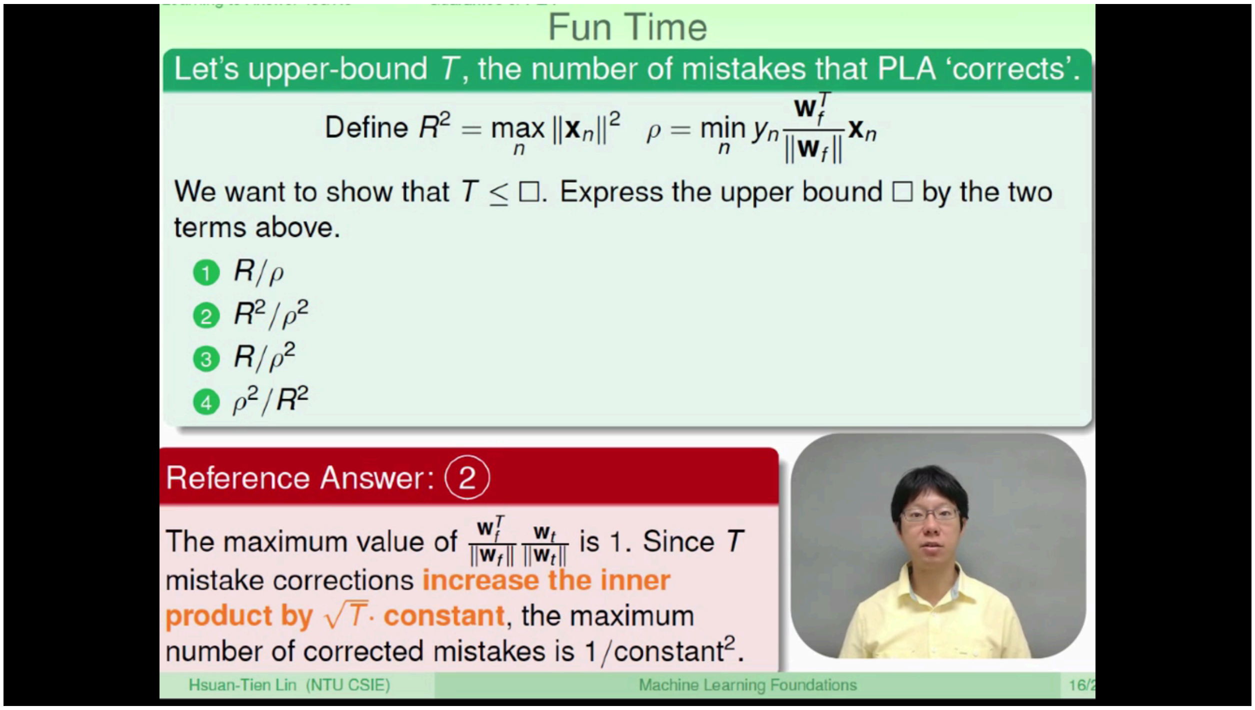
$$\geq \frac{t \cdot \rho}{\sqrt{t} R}$$

$$\geq \sqrt{t} \frac{\rho}{R}$$

所以, 若執行 T 次的 update, 目標線 w_f^T 與預測線 w_t^T 的相似程度, 即內積的結果

$$\frac{w_f^T \cdot w_t}{\|w_f\| \, \|w_t\|} \geq \sqrt{T} \frac{\rho}{R}$$ --------------------(4)

$$\frac{w_f^T \cdot w_t}{\|w_f\| \, \|w_t\|} \geq \sqrt{T} \cdot constant$$ --------------------(5)



計算收斂次數 T 的 upper bound

=> 由 (5) T 次 update W 後, 目標線 w_f^T 與預測線 w_t^T 的單位向量的內積, 如下

$$\frac{w_f^T \cdot w_t}{\|w_f\| \, \|w_t\|} \geq \sqrt{T} \cdot constant$$

而 內積的最大值 = 1, 所以

$$\sqrt{T} \cdot constant \leq 1$$

因此 T 的 upper bound 為

$$T \leq \frac{1}{constant^2}$$

$$\boxed{T \leq \frac{R^2}{\rho^2}}$$

Appendix

計算 T 的 upper bound

⇒ T 次修正 ~~[illegible]~~ 會使內積值 > √T · constant

⇒ 內積的最大值 1

~~內積~~ 最 ~~[illegible]~~ √T constant < 1

⇒ $T \le \frac{1}{constant^2}$

$\|W_{t+1}\|^2 \le \|W_t\|^2 + \max_n \|y_n x_n\|^2$ pp.15

$W_0 = 0$

$W_f^T W_{t+1} \ge W_f^T W_t + \min_n y_n W_f^T x_n$ pp.14

$\rho = \min_n y_n \frac{W_f^T}{\|W_f\|} x_n$

$R^2 = \max_n \|x_n\|^2$

Page 14.

$W_f^T W_0 \to 0$

$W_f^T W_1 \ge \min_n y_n W_f^T x_n$

$W_f^T W_2 \ge W_f^T W_1 + \min_n y_n W_f^T x_n$

$W_f^T W_2 \ge \min_n y_n W_f^T x_n + \min_n y_n W_f^T x_n$

$\ge 2 \min_n y_n W_f^T x_n$

$W_f^T W_3 \ge W_f^T W_2 + \min_n y_n W_f^T x_n$

$\ge 3 \min_n y_n W_f^T x_n$

t次

$W_f^T W_t \ge t \min_n y_n W_f^T x_n$ — ①

$W_f^T W_{t+1} \ge (t+1) \min_n y_n W_f^T x_n$

$W_f^T W_{t+1} \ge (t+1)\, \rho \cdot \|W_f\|$

page 15

$\|W_1\|^2 \le \|W_0\|^2 + \max_n \|y_n x_n\|^2$

$\|W_1\|^2 \le 0 + \max_n \|y_n x_n\|^2$

$\|W_2\|^2 \le \|W_1\|^2 + \max_n \|y_n x_n\|^2$

$\le \max_n \|y_n x_n\|^2 + \max_n \|y_n x_n\|^2$

$\le 2 \max_n \|y_n x_n\|^2$

t次

$\|W_t\|^2 \le t \max_n \|y_n x_n\|^2$ — ②

$\|W_{t+1}\|^2 \le (t+1) \max_n \|y_n x_n\|^2$

$\le (t+1) R^2 \max \|y_n\|^2$

$$\frac{w_f^T}{\|w_f\|} \cdot \frac{w_t}{\|w_t\|} \quad 内積$$

$$\Rightarrow \frac{w_f^T \cdot w_t}{\|w_f\| \cdot \|w_t\|} \xrightarrow{代入①} \geq \frac{t \min_n y_n w_f^T x_n}{\|w_f\| \, \|w_t\|}$$

$$\geq \frac{t \rho \cancel{\|w_f\|}}{\cancel{\|w_f\|} \, \|w_t\|}$$

$$\geq \frac{t \rho}{\|w_t\|}$$

$$\|w_t\|^2 \leq t \max_n \|y_n x_n\|^2$$

$$\frac{1}{\|w_t\|} \geq \sqrt{\frac{1}{t \max_n \|y_n x_n\|^2}}$$

$$\geq \frac{1}{\sqrt{t R^2 \underbrace{\cancel{\max_n \|y_n\|^2}}_{=1}}}$$

$$\frac{t\rho}{\|w_t\|} = \cancel{\frac{t\rho}{tR^2 \max_n \|y_n\|^2}}$$

$$= \frac{t\rho}{\sqrt{t} R}$$

Reference

- https://www.youtube.com/watch?v=Okrrz0IYoSE&list=PLXVfgk9fNX2I7tB6oIINGBmW50rrmFTqf&index=8