

有關 perceptron 收斂問題的推導

--為何 w_f 和 w_t 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 $y_n = \text{sign}(w_f^T x_n)$.

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

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

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

每一次更新 w_{t+1} , 就應該與 w_f 更接近. 因此

$$\begin{aligned} w_f^T w_{t+1} &= w_f^T (w_t + y_{n(t)} x_{n(t)}) \\ &\geq w_f^T w_t + \min_n y_n w_f^T x_n \end{aligned}$$

如 Page 14, 顯示的結果

Learning to Answer Yes/No Guarantee of PLA

PLA Fact: w_t Gets More Aligned with w_f

linear separable $\mathcal{D} \Leftrightarrow$ **exists perfect w_f such that $y_n = \text{sign}(w_f^T x_n)$**

- w_f perfect hence every x_n correctly away from line:
$$y_{n(t)} w_f^T x_{n(t)} \geq \min_n y_n w_f^T x_n > 0$$
- $w_f^T w_t \uparrow$ by updating with any $(x_{n(t)}, y_{n(t)})$
$$\begin{aligned} w_f^T w_{t+1} &= w_f^T (w_t + y_{n(t)} x_{n(t)}) \\ &\geq w_f^T w_t + \min_n y_n w_f^T x_n \\ &> w_f^T w_t + 0. \end{aligned}$$

w_t appears more aligned with w_f after update (really?)

Hsuan-Tien Lin (NTU CSIE) Machine Learning Foundations 14/22

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:

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

w_f 是目標 weight.

=>

t = 0

$$w_0 = 0$$

t = 1

$$w_f^T w_1 \geq w_f^T w_0 + \min_n y_n w_f^T x_n$$

t = 2

$$w_f^T w_2 \geq w_f^T w_1 + \min_n y_n w_f^T x_n$$

$$w_f^T w_2 \geq \min_n y_n w_f^T x_n + \min_n y_n w_f^T x_n$$

$$w_f^T w_2 \geq 2 \cdot \min_n y_n w_f^T x_n$$

t = 3

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

第 t 次

$$w_f^T w_t \geq t \cdot \min_n y_n w_f^T x_n \text{ ----- (1)}$$

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 .

PLA Fact: \mathbf{w}_t Does Not Grow Too Fast \mathbf{w}_t changed only when mistake

$$\Leftrightarrow \text{sign}(\mathbf{w}_t^T \mathbf{x}_{n(t)}) \neq y_{n(t)} \Leftrightarrow y_{n(t)} \mathbf{w}_t^T \mathbf{x}_{n(t)} \leq 0$$

- mistake 'limits' $\|\mathbf{w}_t\|^2$ growth, even when updating with 'longest' \mathbf{x}_n

$$\begin{aligned} \|\mathbf{w}_{t+1}\|^2 &= \|\mathbf{w}_t + y_{n(t)} \mathbf{x}_{n(t)}\|^2 \\ &= \|\mathbf{w}_t\|^2 + 2y_{n(t)} \mathbf{w}_t^T \mathbf{x}_{n(t)} + \|y_{n(t)} \mathbf{x}_{n(t)}\|^2 \\ &\leq \|\mathbf{w}_t\|^2 + 0 + \|y_{n(t)} \mathbf{x}_{n(t)}\|^2 \\ &\leq \|\mathbf{w}_t\|^2 + \max_n \|y_n \mathbf{x}_n\|^2 \end{aligned}$$

start from $\mathbf{w}_0 = \mathbf{0}$, after T mistake corrections,

$$\frac{\mathbf{w}_f^T \mathbf{w}_T}{\|\mathbf{w}_f\| \|\mathbf{w}_T\|} \geq \sqrt{T} \cdot \text{constant}$$

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

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

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

令

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

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

做法如下

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

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

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

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

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

$$\frac{1}{\|w_t\|} \geq \sqrt{\frac{1}{t \cdot \max_n \|y_n x_n\|^2}}, \text{ 因為 } y_n \text{ 不是 } (+1) \text{ 就是 } (-1), \text{ 所以 } \|y_n\|=1.$$

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

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

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

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

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

Fun Time

Let's upper-bound T , the number of mistakes that PLA 'corrects'.

$$\text{Define } R^2 = \max_n \|\mathbf{x}_n\|^2 \quad \rho = \min_n y_n \frac{\mathbf{w}_f^T \mathbf{x}_n}{\|\mathbf{w}_f\|}$$

We want to show that $T \leq \square$. Express the upper bound \square by the two terms above.

- ① R/ρ
- ② R^2/ρ^2
- ③ R/ρ^2
- ④ ρ^2/R^2

Reference Answer: ②

The maximum value of $\frac{\mathbf{w}_f^T \mathbf{w}_t}{\|\mathbf{w}_f\| \|\mathbf{w}_t\|}$ is 1. Since T mistake corrections **increase the inner product by $\sqrt{T} \cdot \text{constant}$** , the maximum number of corrected mistakes is $1/\text{constant}^2$.



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

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

而 內積的最大值 = 1, 所以

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

因此 T 的 upper bound 為

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

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

Appendix

mandatory

description

bucket name



USE FIELDS

constant >

constant <

constant < 1

constant > 1

积

数

内积的最大值

DESCRIPTION

domain_get

MAND NAME

domain_get

CONST 914

积

内积值

会使

下次修正

计算 T 的 upper bound

$$\|W_{t+1}\|^2 \leq \|W_t\|^2 + \max_n \|y_n x_n\|^2 \quad \text{P.P. 15}$$

$W_0 = 0$

$$W_f^T W_{t+1} \geq W_f^T W_t + \min_n y_n W_f^T x_n \quad \text{P.P. 14}$$

①

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

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

Page 14.

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

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

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

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

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

$$\geq 3 \min_n y_n W_f^T x_n \quad \text{①}$$

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

$$W_f^T W_{t+1} \geq (t+1) P \cdot \|W_f\|$$

~~W_0 = 0~~

Page 15

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

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

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

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

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

~~①~~ t 次

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

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

$$\leq (t+1) R^2 \max_n \|y_n\|^2$$

~~W_f^T W_{t+1} \geq (t+1) P \|W_f\|~~

①

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

$$\Rightarrow \frac{w_f^T \cdot w_t}{\|w_f\| \cdot \|w_t\|} \stackrel{\text{Cauchy}}{\geq} \frac{\pm \min_n \lambda_n w_f^T x_n}{\|w_f\| \|w_t\|}$$

$$\geq \frac{\pm p \|w_f\|}{\|w_f\| \|w_t\|}$$

$$\geq \frac{\pm p}{\|w_t\|}$$

$$\|w_t\| \leq \pm \max_n \|x_n\|$$

$$\frac{1}{\|w_t\|} \geq \frac{1}{\pm \max_n \|x_n\|}$$

$$\geq \frac{1}{\sqrt{N} R \max_n \|x_n\|}$$

$$\frac{\pm p}{\|w_t\|} = \frac{\pm p}{\sqrt{N} R \max_n \|x_n\|}$$

$$= \frac{\pm p}{\sqrt{N} R}$$

$$\frac{L}{\sqrt{N} R} = \frac{L}{\sqrt{N} R}$$

Reference

- <https://www.youtube.com/watch?v=Okrrz0IYoSE&list=PLXVfgk9fNX2I7tB6oIINGBmW50rmFTqf&index=8>