$\to 0 \to 1$

Lecture (1) Intro + Perceptron

تبييض لأهم نقاط المحاضرة الأولى

1 The general supervised Learning Algorithm: we have:

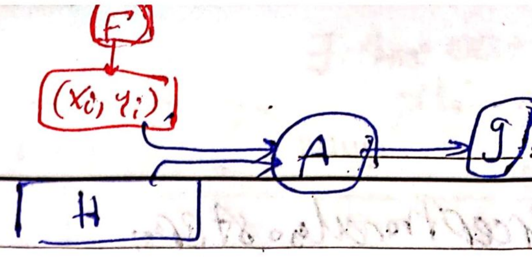
1. Unknown Target function $F : X \to Y$
2. Training Data $(X_1, Y_1), (X_2, Y_2), \ldots, (X_N, Y_N)$ where $Y_i = f(X_i)$
3. want to learn g close to F
4. To learn $g \to$ Learning Models
5. To say how close g is to $F \to$ Learning Theory.

2 How to learn F?

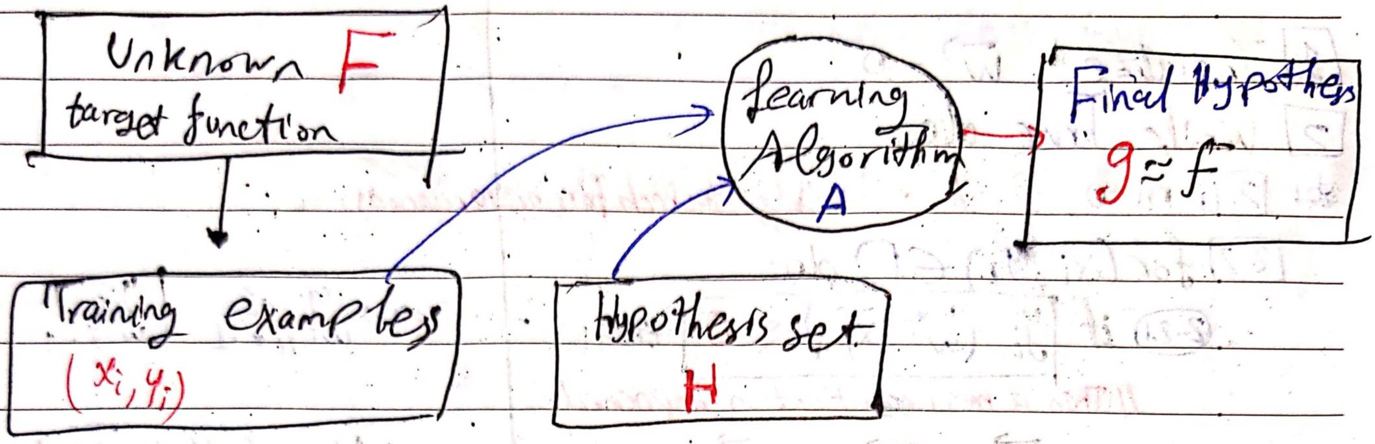
1. Pick a Hypothesis set $H = \{h_1, h_2, \ldots\}$ → set of all possible values
2. Use a learning algorithm to select a hypothesis g from H based on the training data

- H + learning algorithm = learning model.

اختيارهم مرتبط ببعض جدا.



3 Learning setup:



4 Basic classification example: we have two classes: yes / no

→ yes if $\sum_{i=1}^{d} w_i x_i \geq$ threshold, no if $\sum_{i=1}^{d} w_i x_i <$ threshold. [linear separation]

∴ linear formula $h \in H$:

$$h(x) = \text{sign}\left(\sum_{i=1}^{d} w_i x_i - \text{threshold}\right) = \text{sign}\left(\sum_{i=1}^{d} w_i x_i + w_0\right)$$

$$h(x) = \text{sign}\left(\sum_{i=0}^{d} w_i x_i\right) \quad \text{for } x_0 = 1$$

5 The Perceptron Algorithm:

$$h(x) = \text{sign}\left(\sum_{i=0}^{d} w_i x_i\right) \xrightarrow[\text{form}]{\text{vector}} \boxed{h(x) = \text{sign}\left(w^T X\right)}$$

Prediction

Perceptron Algo assumptions:

1. Binary classification $y = \{+1, -1\}$
2. Linearly seperable data

الموضوع: ______ التاريخ: / /

$\boxed{6}$ Perceptron algo steps:

$\boxed{1}$ Initialize $\vec{w} = 0$

$\boxed{2}$ While True do:

$\boxed{2.1}$ $m=0$ // #changes (to check for convergance)

$\boxed{2.2}$ for $(x_i, y_i) \in D$ do:

(2.2.1) if $\boxed{y_i \, (\vec{w}^T \cdot \vec{x}_i) \leq 0}$ then: → Cases i.f $y_i = +1$, $w^T x_i = -1$ | $y_i = -1$, $w^T x = 1$

// Then a misclassification happened.

2.2.1.1: $\vec{w} \leftarrow \vec{w} + y\vec{x}$ //update weights in the right direction of correctly classifying $x(t)$

2.2.1.2: $m \leftarrow m+1$

$\boxed{2.3}$ if $m=0$: // no changes happened. (converged)

(2.3.1) break

$\boxed{7}$ proving that w moves at right direction:

~~after we do update 2.2.2.1 $\vec{w} \leftarrow \vec{w} + y\vec{x}$~~ at step $t+1$ iteration (a misclass. happened)

$$\underset{\leq 0}{y_t} \, w_{t+1}^T x_t = y_t [w_t + y_t x_t]^T x_t$$

$$= y_t w_t^T x_t + \underbrace{y_t^2}_{1 \text{ since } y \in \{-1,1\}} x_t^T x_t$$

$$= \underline{y_t w_t^T x_t + \|x_t\|^2} \geq y_t w_t^T x_t$$

يعني بتزيد القيمة دي وتبعد عن كونها سالب / فالسالب يقل وتقترب من انها

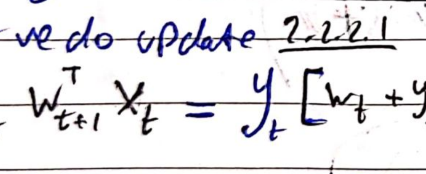
تتصل 2 وتبقى ≥ 0

(8) Cases of misclassification

$y = +1$	$w^T \cdot x \leq 0$	$y = -1$	$w^T \cdot x > 0$

$y = +1$:

الاتنين بيشاوروا في اتجاهات مختلفة (θ منفرجة)

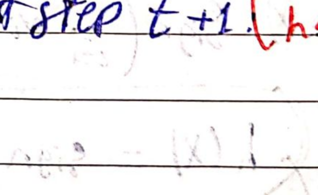


$W_{t+1} = W_t + 1 x_t$

عشان ال dot product يكون $y w^T x > 0$ الزاوية الجاية تبقى حادة

$y = -1$:

بينهما زاوية حادة عاوزين نبعدهم عن بعض عشان المرة اللي بعدها ال dot product يكون بالسالب، ويكون $y w^T x > 0$ correct classification



∃ : there exist
يعني يوجد في واحد كدة

التاريخ: / /

9 Perceptron Convergance proof

الموضوع:

initialization:
* Data is linearly separable
* therefore: ∃ w^* that separates the training data
* assume @ $t=0$ $\quad w_t \Rightarrow w_0 = 0$

Required: to prove that perceptron algorithm finds a linear separator vector (w) after a finite number of iterations t

Derivation of t :-

1 let $\rho = \min_n y_n (w^{*T} x_n)$

$\rho > 0$? because w^* perfectly separates data points : correct prediction

→ same signs for both y_n & $w^{*T} x_n$ ✓ > 0

2 we know that $\quad w_t = w_{t-1} + y_{t-1} x_{t-1}$

transpose → dot product with w^* →

$$\underbrace{w_t^T w^*}_{1} = \underbrace{w_{t-1}^T w^*}_{2} + y_{t-1} x_{t-1}^T w^* \geq w_{t-1}^T w^* + \rho$$

أكبر منه عشان ال ρ هي أصغر قيمة لل $y_n w^{*T} x_n$

عرفت الترم 2 ؟ هو نفس 1 بس في $(t-1)$ الخطوة اللي قبلها

جربت تعوض كدة:

$$w_t^T w^* = w_{t-2}^T w^* + y_{t-2} x_{t-2}^T w^* + y_{t-1} x_{t-1}^T w^*$$

عوضنا عن 2 ودة كدة w

دة أصغر من ال ρ أو يساويه

بص كدة

$$\Rightarrow w_t^T w^* \geq w_{t-2}^T w^* + \rho + \rho$$

$$w_t^T w^* \geq w_{t-2}^T w^* + 2\rho$$

لو فضلت ماشي recursive

$t \Leftarrow$ iteration بعدكلل

$$w_t^T w^* \geq w_{t-t}^T w^* + t\rho$$

$w_0^T = 0 \rightarrow$ given!

$$\boxed{\therefore w_t^T w^* \geq t\rho}$$

مرده ليه t بالظبط!

عشان انت بعد أول iteration قولت انه أكبر من ρ واحدة

بعد كمان واحدة قولت انه أكبر من 2ρ صح!

$t\rho$ بعد t ؟ ← هتقول انه أكبر من

3 Note: $\|a+b\|^2 = \|a\|^2 + 2a^T b + \|b\|^2$

where $\|v\|$ is the norm of vector v

we know that:

$$W_t = W_{t-1} + X_{t-1} y_{t-1} \xrightarrow{\text{take norm of both sides}}$$

$$\|W_t\|^2 = \|W_{t-1}\|^2 + \|X_{t-1} y_{t-1}\|^2$$

لعمته   لعمته   لبرة

$$\therefore \|W_t\|^2 = \|W_{t-1}\|^2 + \underbrace{\|y_{t-1}\|^2}_{1} \|X_{t-1}\|^2 + \underbrace{2 y_{t-1} W_{t-1}^T X_{t-1}}_{\text{الرقم ده أقل من صفر ، ليه؟}}$$

⟵ عشان هو جايالنا بنحسب W_t يعني بنعمل weight update صح؟ ودة بيحصل في حالة حصل misclassification بس ، اللي هو لما حاصل ضرب $y W^T x$ يكون سالب .

$$\therefore \|W_t\|^2 \leq \|W_{t-1}\|^2 + \|X_{t-1}\|^2$$

$$R = \max_n \|X_n\|$$ ⟵ لو عرفنا متغير جديد اسمه R حنقسمه *

ومشينا بال recursion زي المرة اللي فاتت وعوضنا عن الترم ده هنوصل ل .

$$\boxed{\|W_t\|^2 \leq t R^2}$$

لو مش متخيلها بص كده :

$$\therefore \|W_t\|^2 \leq \|W_{t-1}\|^2 + \|X_{t-1}\|^2 \leq \|W_{t-1}\|^2 + R^2$$

$$\& \ \|W_{t-1}\|^2 \leq \|W_{t-2}\|^2 + \|X_{t-1}\|^2 \leq \|W_{t-2}\|^2 + R^2$$

جمع كده الاتنين وعوض عن X بـ انها أقل برده من الأكتر

$$\therefore \|W_t\|^2 \leq \|W_{t-2}\|^2 + R^2 + R^2$$

$$\therefore \|W_t\|^2 \leq \|W_{t-2}\|^2 + 2R^2 \longrightarrow \text{after 2 iteration}$$

$$\therefore \|W_t\|^2 \leq \|W_{t-t}\|^2 + tR^2$$

كمل بنفس النمط هتوصل ل t من المرات

$$\|W_0\|^2 = 0$$

$$\therefore \|W_t\|^2 \leq t R^2$$

بعد الحمد لله ...

$\boxed{4}$ we know that dot product rule is $a.b = \|a\|\,\|b\|\cos\theta$

وكما رأينا إن الـ $\cos\theta$ بيتراوح بين الـ -1 والـ (1) صح؟ يعني أكبر قيمة للـ $\cos\theta = 1$ عوضوه

يعني الـ $\|a\|\|b\|$ وهي أكبر قيمة للـ dot Product $a.b$

بس وبيحصل في حالة الـ $\cos\theta = 1$ ، $\|a\|\,\|b\| \geq \|a\|\,\|b\|\cos\theta$

$$\boxed{w_t^T \cdot w^* = \|w_t\|\,\|w^*\|\cos\theta \leq \|w_t\|\,\|w^*\|}$$ نعمل تشابه بقاعدة الـ a.b

$\therefore$ $w_t^T w^* \leq \|w_t\|\,\|w^*\|$

• استنتجنا من خطوة $\boxed{2}$ أن

From $\boxed{2}$ $w_t^T w^* \geq t\rho$

$\therefore$ $\|w_t\|\,\|w^*\| \geq t\rho \rightarrow$ (I)

• " " " $\boxed{3}$ أن

From $\boxed{3}$: $\|w_t\|^2 \leq tR^2 \rightarrow \boxed{\|w_t\| \leq \sqrt{t}R} \rightarrow$ (II)

from (I) & (II)

$\sqrt{t}R \geq \|w_t\| \geq \dfrac{t\rho}{\|w^*\|}$ $\therefore$ $\cancel{\sqrt{t}}R \geq \dfrac{\cancel{t}^{\sqrt{t}}\rho}{\|w^*\|}$

$$\boxed{\therefore\ t \leq \frac{R^2\,\|w^*\|^2}{\rho^2}}$$

وصلنا الحمدلله :)

وصلنا لإيه؟

احنا كنا عاوزين نثبت إيه تاني (:)

$\therefore$ آه إن الـ Perceptron بت converge ، أي إنها لا تستمر إلى ما لا نهاية

$\therefore t$ (number of iterations) is finite, has an upper bound

• طب هنعرف نحسبه؟ أكيد لا عشان هو محتاج w^* ولو معايا w^* هرمي الـ Perceptron كله جنبه أصلاً XD

← الغرض بس من القانون ده إنه يطمنا إننا هنوصل إن شاء الله.

Notes (1) w^* is any linear separator, which we don't know previously

(2) we don't say that w_t converges to w^*, we just say that w_t converges to a linear separator.

مش هنعرف نستخدم الـ Perceptron مع الـ XOR Problem عشان مش linearly separable

---

Lecture 1 is Done :)

تمت بحمد الله