

Appunti di Programmazione Matematica

Canale 1 · Prof. Marco Sciandrone

Capitolo 10: Il neurone artificiale

Autore: Tommaso C.

Appunti tratti dalle lezioni in aula e dal materiale del corso

Argomenti: Perceptron · Classificatore lineare · Addestramento
Funzione logistica · XOR e reti di neuroni

Indice

1	Il Perceptron	1
2	Un classificatore lineare	1
3	Addestrare un neurone	2
4	Una funzione di errore derivabile	3
5	Lo XOR e le reti di neuroni	4
6	Esercizi	5
7	Soluzioni	6

1 Il Perceptron

Definizione – Perceptron

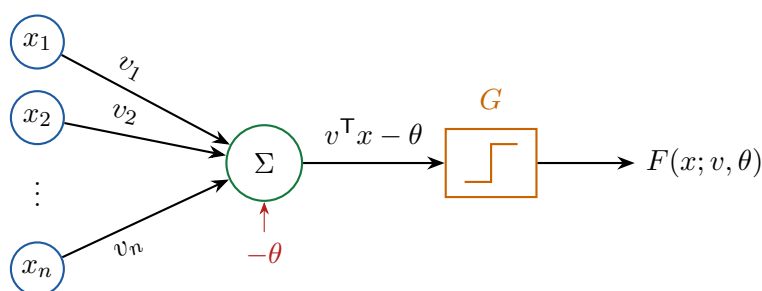
Il **Perceptron** è il modello di neurone con ingressi $x = (x_1, \dots, x_n)$, **pesi** $v = (v_1, \dots, v_n)$ e **soglia** $\theta \in \mathbb{R}$, la cui uscita è

$$F(x; v, \theta) = G(v^\top x - \theta) = G\left(\sum_{i=1}^n v_i x_i - \theta\right),$$

dove G è la **funzione di attivazione** a gradino:

$$G(t) = \begin{cases} 1 & \text{se } t \geq 0, \\ 0 & \text{se } t < 0. \end{cases}$$

Il modello imita il neurone biologico: riceve segnali da altri neuroni, ciascuno moltiplicato per un peso sinaptico, e si attiva quando la somma pesata raggiunge la soglia. Il neurone vale 1 (attivo) se $\sum_i v_i x_i \geq \theta$, 0 (inattivo) altrimenti.



Il Perceptron: somma pesata degli ingressi, meno la soglia, passata attraverso il gradino G .

2 Un classificatore lineare

Proposizione – Il neurone divide il piano con una retta

Con $n = 2$ e $(v_1, v_2) \neq (0, 0)$, l'insieme

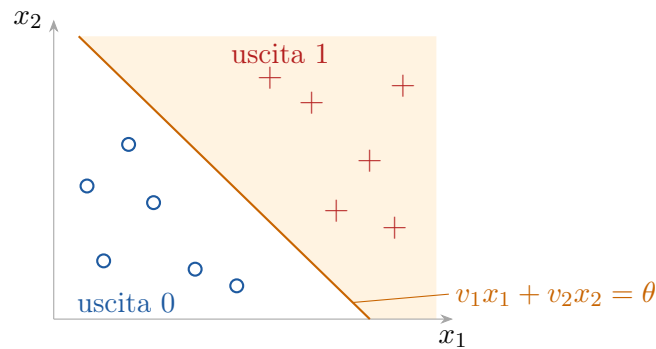
$$\{(x_1, x_2) \in \mathbb{R}^2 : v_1 x_1 + v_2 x_2 = \theta\}$$

è una retta. Il neurone dà uscita 1 nel semipiano $v_1 x_1 + v_2 x_2 \geq \theta$ e uscita 0 nel semipiano $v_1 x_1 + v_2 x_2 < \theta$.

Per questo il Perceptron è un **classificatore lineare**: separa gli ingressi in due classi con una superficie di separazione lineare, la retta per $n = 2$ e l'insieme $\{x \in \mathbb{R}^n : v^\top x = \theta\}$ in generale. Pesi e soglia decidono dove sta la retta.

Esempio – Pazienti a rischio

Di ogni paziente si misurano due caratteristiche, pressione arteriosa x_1 e fluidità del sangue x_2 : il paziente è un punto di $\mathbb{R}^2$. Dopo cinque anni si assegna l'etichetta 1 a chi ha avuto un evento ischemico e 0 agli altri. Una retta che separa i due gruppi permette di classificare un paziente nuovo dal semipiano in cui cade.



Pazienti con etichetta 1 (croci) e 0 (cerchi): la retta lascia le croci nel semipiano dove il neurone vale 1.

3 Addestrare un neurone

Definizione – Training set

Il **training set** è un insieme di P coppie di dati

$$\mathcal{T} = \{(x^{(p)}, y^{(p)}) : p = 1, \dots, P\},$$

dove $x^{(p)} \in \mathbb{R}^n$ è il vettore delle caratteristiche e $y^{(p)} \in \{0, 1\}$ la sua etichetta nota.

Addestrare il neurone significa scegliere v e θ in modo che classifichi correttamente i dati noti:

$$F(x^{(p)}; v, \theta) = y^{(p)}, \quad p = 1, \dots, P.$$

I dati $(x^{(p)}, y^{(p)})$ sono parametri assegnati; le variabili sono i pesi e la soglia, $n + 1$ numeri.

Le P equazioni si trasformano in un problema di minimo. Si misura l'errore sul campione p

$$e_p(v, \theta) = |F(x^{(p)}; v, \theta) - y^{(p)}|,$$

che è nullo solo se il neurone classifica bene $x^{(p)}$, e si somma su tutti i campioni:

$$\min_{v \in \mathbb{R}^n, \theta \in \mathbb{R}} \sum_{p=1}^P e_p(v, \theta).$$

Il problema si risolve con un algoritmo, come il metodo del gradiente (Capitolo 8), che usa le derivate della funzione da minimizzare.

4 Una funzione di errore derivabile

L'errore così scritto non è continuo. Fissato un campione con $y^{(p)} = 1$, si guardi e_p come funzione della soglia θ , con $s = v^T x^{(p)}$:

$$e_p = \begin{cases} 0 & \text{se } \theta \leq s, \\ 1 & \text{se } \theta > s. \end{cases}$$

L'errore salta da 0 a 1 in $\theta = s$, come il gradino G in 0. Un algoritmo che usa le derivate non si può applicare.

Le due cause si correggono una alla volta.

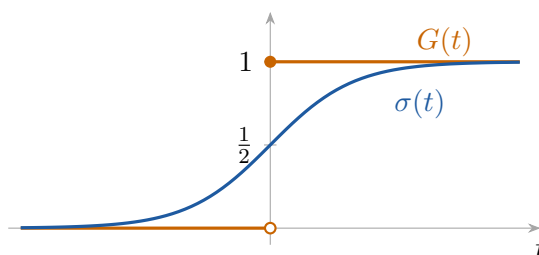
1. **Attivazione.** Il gradino si sostituisce con la funzione logistica σ (definizione sotto), che ha andamento simile ma è continua e derivabile.
2. **Errore.** Il valore assoluto $|t|$ non è derivabile in $t = 0$: si sostituisce con il quadrato t^2 , che come $|t|$ è nullo solo per $t = 0$ e cresce con $|t|$.

Definizione – Funzione logistica

La **funzione logistica** è

$$\sigma(t) = \frac{1}{1 + e^{-t}}, \quad t \in \mathbb{R}.$$

σ è continua e derivabile su $\mathbb{R}$, con valori in $(0, 1)$, $\sigma(0) = \frac{1}{2}$, $\sigma(t) \rightarrow 0$ per $t \rightarrow -\infty$ e $\sigma(t) \rightarrow 1$ per $t \rightarrow +\infty$.



Il gradino G salta in 0; la logistica σ passa da 0 a 1 senza salti.

Con queste due sostituzioni il neurone diventa $F(x; v, \theta) = \sigma(v^T x - \theta)$ e il problema di addestramento è

$$\min_{v \in \mathbb{R}^n, \theta \in \mathbb{R}} \sum_{p=1}^P \left(\sigma(v^T x^{(p)} - \theta) - y^{(p)} \right)^2$$

La funzione da minimizzare è continua e derivabile rispetto a ciascun peso v_i e alla soglia θ , tenendo fisse le altre variabili: un algoritmo con le derivate si può applicare.

5 Lo XOR e le reti di neuroni

Lo **XOR** (o esclusivo) assegna a due ingressi binari l'etichetta 1 se sono diversi e 0 se sono uguali:

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

Proposizione – Un neurone non basta

Non esistono v_1, v_2, θ tali che il Perceptron classifichi correttamente i quattro punti dello XOR.

Dimostrazione

Se esistessero, da $G(t) = 1 \iff t \geq 0$:

1. da $(0, 0) \mapsto 0$: $-\theta < 0$, cioè $\theta > 0$;
2. da $(1, 0) \mapsto 1$ e $(0, 1) \mapsto 1$: $v_1 \geq \theta$ e $v_2 \geq \theta$, quindi $v_1 + v_2 - \theta \geq \theta > 0$;
3. da $(1, 1) \mapsto 0$: $v_1 + v_2 - \theta < 0$, contro il passo 2.

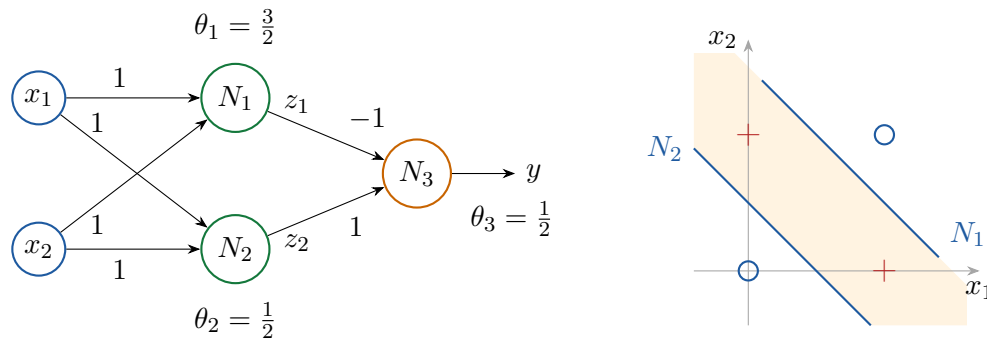
□

In termini geometrici, nessuna retta lascia $(0, 1)$ e $(1, 0)$ da una parte e $(0, 0)$ e $(1, 1)$ dall'altra.

Una rete di tre neuroni risolve il problema. I neuroni N_1 e N_2 ricevono x_1, x_2 ; le loro uscite z_1, z_2 sono gli ingressi del neurone N_3 , che dà l'uscita della rete:

$$z_1 = G\left(x_1 + x_2 - \frac{3}{2}\right), \quad z_2 = G\left(x_1 + x_2 - \frac{1}{2}\right), \quad y = G\left(-z_1 + z_2 - \frac{1}{2}\right).$$

Per l'ingresso $(0, 1)$: $x_1 + x_2 = 1$, quindi $z_1 = G(-\frac{1}{2}) = 0$, $z_2 = G(\frac{1}{2}) = 1$ e $y = G(\frac{1}{2}) = 1$. Gli altri tre ingressi sono l'Esercizio 1.



A sinistra la rete, con i pesi sulle frecce e le soglie. A destra le rette di N_2 ($x_1 + x_2 = \frac{1}{2}$) e di N_1 ($x_1 + x_2 = \frac{3}{2}$): la rete vale 1 nella striscia fra le due, dove stanno $(0,1)$ e $(1,0)$.

I pesi e le soglie della rete sono 9 numeri, 3 per neurone. Qui sono stati scelti a mano; in pratica li sceglie l'addestramento, minimizzando come nella Sezione 4 una funzione di errore di 9 variabili sui dati dello XOR.

Riepilogo

- **Perceptron:** $F(x; v, \theta) = G(v^\top x - \theta)$, con $G(t) = 1$ per $t \geq 0$ e 0 per $t < 0$.
- **Classificatore lineare:** uscita 1 nel semipiano $v_1 x_1 + v_2 x_2 \geq \theta$, 0 nell'altro.
- **Training set:** $\mathcal{T} = \{(x^{(p)}, y^{(p)})\}$, con $y^{(p)} \in \{0, 1\}$.
- **Addestramento:** scegliere v, θ con $F(x^{(p)}; v, \theta) = y^{(p)}$, minimizzando la somma degli errori.
- **Regolarità:** il gradino e $|t|$ rendono l'errore discontinuo o non derivabile.
- **Logistica:** $\sigma(t) = 1/(1 + e^{-t})$, continua e derivabile.
- **Errore derivabile:** $\min_{v, \theta} \sum_p (\sigma(v^\top x^{(p)} - \theta) - y^{(p)})^2$.
- **XOR:** un neurone non basta; una rete di tre neuroni (9 parametri) sì.

6 Esercizi

Esercizio 1 – La rete dello XOR

Verificare che la rete di tre neuroni della Sezione 5,

$$z_1 = G\left(x_1 + x_2 - \frac{3}{2}\right), \quad z_2 = G\left(x_1 + x_2 - \frac{1}{2}\right), \quad y = G\left(-z_1 + z_2 - \frac{1}{2}\right),$$

dà l'uscita dello XOR sui quattro ingressi $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$.

Soluzione a pagina 6 →

7 Soluzioni

Soluzione dell'Esercizio 1 – La rete dello XOR

z_1 e z_2 dipendono solo da $s = x_1 + x_2$: $z_1 = 1$ se e solo se $s \geq \frac{3}{2}$, $z_2 = 1$ se e solo se $s \geq \frac{1}{2}$.

(x_1, x_2)	s	z_1	z_2	$-z_1 + z_2 - \frac{1}{2}$	y
(0, 0)	0	0	0	$-\frac{1}{2}$	0
(0, 1)	1	0	1	$\frac{1}{2}$	1
(1, 0)	1	0	1	$\frac{1}{2}$	1
(1, 1)	2	1	1	$-\frac{1}{2}$	0

La colonna y coincide con l'etichetta dello XOR in tutti e quattro i casi. N_1 vale 1 solo se entrambi gli ingressi sono 1, N_2 se almeno uno lo è; N_3 vale 1 quando N_2 è attivo e N_1 no.

[← Torna all'Esercizio 1, pagina 5](#)