

Appunti di Programmazione Matematica

Canale 1 · Prof. Marco Sciandrone

Capitolo 8: Il metodo del gradiente

Autore: Tommaso C.

Appunti tratti dalle lezioni in aula e dal materiale del corso

Argomenti: Minimizzazione in una variabile · Direzione opposta alla derivata
Scelta del passo · Dimezzamento

Indice

1	Il problema e il metodo	1
2	La scelta del passo	1

1 Il problema e il metodo

Sia $f : \mathbb{R} \rightarrow \mathbb{R}$ di classe C^1 , cioè derivabile con derivata f' continua (una funzione *smooth*). Si cerca un punto di minimo di f , ammesso che esista, con un algoritmo: una regola che genera punti $x_0, x_1, x_2, \dots$ in cui f diminuisce.

Definizione – Metodo del gradiente

Dato un punto iniziale x_0 , il **metodo del gradiente** calcola, per $k = 0, 1, 2, \dots$ e finché $f'(x_k) \neq 0$,

$$x_{k+1} = x_k - \alpha_k f'(x_k),$$

con un **passo** $\alpha_k > 0$ scelto in modo che $f(x_{k+1}) < f(x_k)$.

Il metodo genera una successione $\{x_k\}$ (Capitolo 5) con valori $f(x_k)$ strettamente decrescenti. Si vuole che converga a un punto di minimo.

La regola nasce dall'approssimazione lineare del Capitolo 7, $\hat{f}(x) = f(x_k) + f'(x_k)(x - x_k)$, che in x_k vale $f(x_k)$:

1. $\hat{f}(x) < f(x_k)$ se e solo se $f'(x_k)(x - x_k) < 0$;
2. se $f'(x_k) > 0$ serve $x < x_k$, se $f'(x_k) < 0$ serve $x > x_k$: ci si sposta in verso opposto al segno della derivata;
3. le due regole insieme danno $x = x_k - \alpha f'(x_k)$ con $\alpha > 0$.

Lungo questa direzione $\hat{f}$ diminuisce per ogni $\alpha > 0$; f non è lineare, e diminuisce solo se α è abbastanza piccolo.

2 La scelta del passo

Proposizione – Un passo piccolo fa diminuire f

Sia f derivabile in x_k con $f'(x_k) \neq 0$. Allora esiste $\bar{\alpha} > 0$ tale che

$$f(x_k - \alpha f'(x_k)) < f(x_k) \quad \text{per ogni } \alpha \in (0, \bar{\alpha}).$$

Dimostrazione

Sia $g = f'(x_k) \neq 0$.

1. Per la definizione di derivata con incremento $h = -\alpha g$, che tende a 0 per $\alpha \rightarrow 0^+$,

$$\frac{f(x_k - \alpha g) - f(x_k)}{\alpha} = -g \cdot \frac{f(x_k - \alpha g) - f(x_k)}{-\alpha g} \longrightarrow -g^2 < 0.$$

2. Per la permanenza del segno esiste $\bar{\alpha} > 0$ tale che il rapporto è negativo per $0 < \alpha < \bar{\alpha}$.
3. Poiché $\alpha > 0$, è negativo il numeratore: $f(x_k - \alpha g) < f(x_k)$.

□

La proposizione non dice quanto vale $\bar{\alpha}$. Il passo si cerca per tentativi, iterazione per iterazione:

1. si prova $\alpha = 1$;
2. se $f(x_k - \alpha f'(x_k)) < f(x_k)$ si accetta: $\alpha_k = \alpha$;
3. altrimenti si dimezza α e si torna al punto 2.

I passi provati sono $1, \frac{1}{2}, \frac{1}{4}, \dots$: dopo un numero finito di dimezzamenti $\alpha < \bar{\alpha}$, e il passo viene accettato.

Esempio – Il minimo di $x^4/4 - x$

Con $f(x) = \frac{1}{4}x^4 - x$, $f'(x) = x^3 - 1$ e $x_0 = 2$:

k	x_k	$f(x_k)$	$f'(x_k)$	α_k	x_{k+1}
0	2	2	7	1/4	0,25
1	0,25	-0,2490	-0,9844	1	1,2344
2	1,2344	-0,6540	0,8808	1/2	0,7940
3	0,7940	-0,6946	-0,4995	1/2	1,0437

Per $k = 0$ i passi 1 e $\frac{1}{2}$ danno -5 e $-1,5$, dove f vale 161,25 e 2,77: non meno di $f(2) = 2$. Il punto di minimo è $\bar{x} = 1$, dove $f' = 0$ e $f(1) = -0,75$; i punti si alternano ai due lati di $\bar{x}$, con $f(x_4) = -0,7470$.

Osservazione – Learning rate e backpropagation

Nell'addestramento delle reti neurali il passo α_k si chiama **learning rate** (tasso di apprendimento). La **backpropagation** non è il metodo di minimizzazione: è il nome delle formule ricorsive che calcolano le derivate prime della funzione di una rete neurale, usate poi dal metodo del gradiente.

Riepilogo

- **Problema:** minimo di $f : \mathbb{R} \rightarrow \mathbb{R}$ di classe C^1 con un algoritmo.
- **Direzione:** verso opposto al segno di $f'(x_k)$, dove l'approssimazione lineare diminuisce.
- **Metodo:** $x_{k+1} = x_k - \alpha_k f'(x_k)$, $\alpha_k > 0$, finché $f'(x_k) \neq 0$.
- **Passo piccolo:** se $f'(x_k) \neq 0$, per $\alpha \in (0, \bar{\alpha})$ il valore di f diminuisce.
- **Dimezzamento:** si prova $\alpha = 1, \frac{1}{2}, \frac{1}{4}, \dots$ fino alla prima diminuzione.
- **Nomi:** il passo è il learning rate; la backpropagation calcola le derivate.