3.1 Metodo di discesa del gradiente

Consideriamo il problema di minimizzare una funzione convessa differenziabile $g : \mathbb{R}^N \to \mathbb{R}$. Indichiamo con w^* un minimo globale di g in $\mathbb{R}^N$. Assumiamo di saper calcolare il gradiente di g in w per ogni $w \in \mathbb{R}^N$ dato. Per la convessità di g, il gradiente $\nabla g(w)$ soddisfa

$$g(w) - g(z) \leq \nabla g(w)^\top (w - z) \qquad \text{per ogni } z \in \mathbb{R}^N. \tag{3.1}$$

Assumiamo inoltre l'esistenza di $D, G > 0$ tali che $\|\nabla g(w)\| \leq G$ per ogni $w \in \mathbb{R}^N$, e $\left\|w^{(1)} - w^*\right\| \leq D$ per qualche $w^{(1)} \in \mathbb{R}^N$. Il punto $w^{(1)}$ sarà il punto iniziale del metodo.

Sia $\eta > 0$ un parametro. La regola di aggiornamento del metodo di discesa del gradiente è la seguente:

$$\boxed{w^{(t+1)} = w^{(t)} - \eta \, \nabla g(w^{(t)}).} \tag{3.2}$$

L'algoritmo di discesa del gradiente applica la regola (3.2) per un certo numero di passi T e restituisce infine il vettore dal valore minimo tra tutti quelli generati:

$$\boxed{w^{GD} \overset{\text{def}}{=} \operatorname*{argmin}_{t=1,2,\dots,T} g(w^{(t)}).}$$

Teorema 3.1.1. *Il metodo di discesa del gradiente (3.2) con $\eta = \frac{D}{G\sqrt{T}}$ soddisfa*

$$\frac{1}{T} \sum_{t=1}^{T} \left(g(w^{(t)}) - g(w^*) \right) \leq \frac{DG}{\sqrt{T}}.$$

In particolare, si ha $g(w^{GD}) \leq g(w^*) + DG/\sqrt{T}$.

Dimostrazione. Per la proprietà (3.1) e la definizione (3.2),

$$\begin{aligned} g(w^{(t)}) - g(w^*) &\leq \nabla g(w^{(t)})^\top (w^{(t)} - w^*) \\ &= \frac{1}{\eta} (w^{(t)} - w^{(t+1)})^\top (w^{(t)} - w^*). \end{aligned}$$

Poiché per ogni coppia di vettori a e b, $2a^\top b = \|a\|^2 + \|b\|^2 - \|a - b\|^2$,

$$\begin{aligned} \frac{1}{\eta}(w^{(t)} - w^{(t+1)})^\top (w^{(t)} - w^*) &= \frac{1}{2\eta} \left(\left\|w^{(t)} - w^{(t+1)}\right\|^2 + \left\|w^{(t)} - w^*\right\|^2 - \left\|w^{(t+1)} - w^*\right\|^2 \right) \\ &= \frac{\eta}{2} \left\|\nabla g(w^{(t)})\right\|^2 + \frac{1}{2\eta} \left(\left\|w^{(t)} - w^*\right\|^2 - \left\|w^{(t+1)} - w^*\right\|^2 \right) \end{aligned}$$

dove la seconda uguaglianza segue dalla (3.2). Usando la definizione di G, possiamo maggiorare l'ultima espressione con

$$\frac{\eta G^2}{2} + \frac{1}{2\eta} \left(\left\|w^{(t)} - w^*\right\|^2 - \left\|w^{(t+1)} - w^*\right\|^2 \right).$$

Mediando su $t = 1, 2, \dots, T$, e ricordando che $\eta \overset{\text{def}}{=} D/G\sqrt{T}$,

$$\frac{1}{T} \sum_{t=1}^{T} \left(g(w^{(t)}) - g(w^*) \right) \leq \frac{\eta G^2 T}{2T} + \frac{1}{2\eta T} \left\|w^{(1)} - w^*\right\|^2$$

$$
\begin{aligned}
&\leq \frac{\eta G^2}{2} + \frac{D^2}{2\eta T} \\
&= \frac{DG}{\sqrt{T}}.
\end{aligned}
$$

La seconda parte segue dal fatto che il minimo di una successione finita non è maggiore della media della stessa successione. □