Algoritmo GD: $w^{(t+1)} = w^{(t)} - \eta \nabla g(w^{(t)})$ (GD)

RETURN $\operatorname{argmin}_{t=1}^{T} g(w^{(t)})$

Assunzioni: (1) $\| w^{(1)} - w^* \| \le D$

(2) $\| \nabla g(w^{(t)}) \| \le G$

(3) $\eta = \dfrac{D}{G\sqrt{T}}$

Tesi: $g(w^{GD}) - g(w^*) \le DG/\sqrt{T}$

$$w^{(1)} \xleftrightarrow{\;\le D\;} w^*$$

(GD) $\nearrow \; = w^{(t)} - \eta \nabla g(w^{(t)})$

$$\|w^{(t+1)} - w^*\|^2 =$$

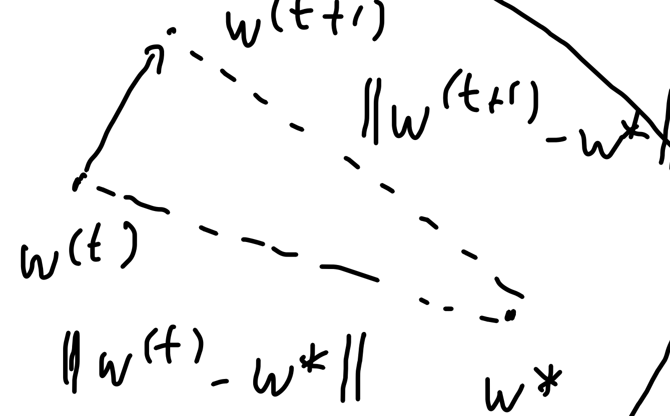
$$= \| w^{(t)} - \eta \nabla g(w^{(t)}) - w^* \|^2$$

$$= \langle w^{(t)} - \eta \nabla g(w^{(t)}) - w^*, \; w^{(t)} - \eta \nabla g(w^{(t)}) - w^* \rangle$$

$$= \langle \overbrace{w^{(t)} - w^*}^{a} \; \overbrace{- \eta \nabla g(w^{(t)})}^{b}, \; \overbrace{w^{(t)} - w^*}^{a} \; \overbrace{- \eta \nabla g(w^{(t)})}^{b} \rangle$$

$$\overset{\downarrow}{=} \|w^{(t)} - w^*\|^2 - 2\eta \langle \nabla g(w^{(t)}), w^{(t)} - w^* \rangle$$

$$+ \eta^2 \underbrace{\|\nabla g(w^{(t)})\|^2}_{\le G^2}$$



$$\|z\|^2 = z^T z$$

$$= \langle z, z \rangle$$

$$\langle a + b, c + d \rangle$$

$$= \langle a, c \rangle + \langle b, c \rangle + \langle b, d \rangle + \langle a, d \rangle$$

$$\|w^{(t+1)} - w^*\|^2 \leq \|w^{(t)} - w^*\|^2 - 2\langle \eta \nabla g(w^{(t)}), w^{(t)} - w^* \rangle + \eta^2 G^2$$

$$\frac{1}{2\eta}$$

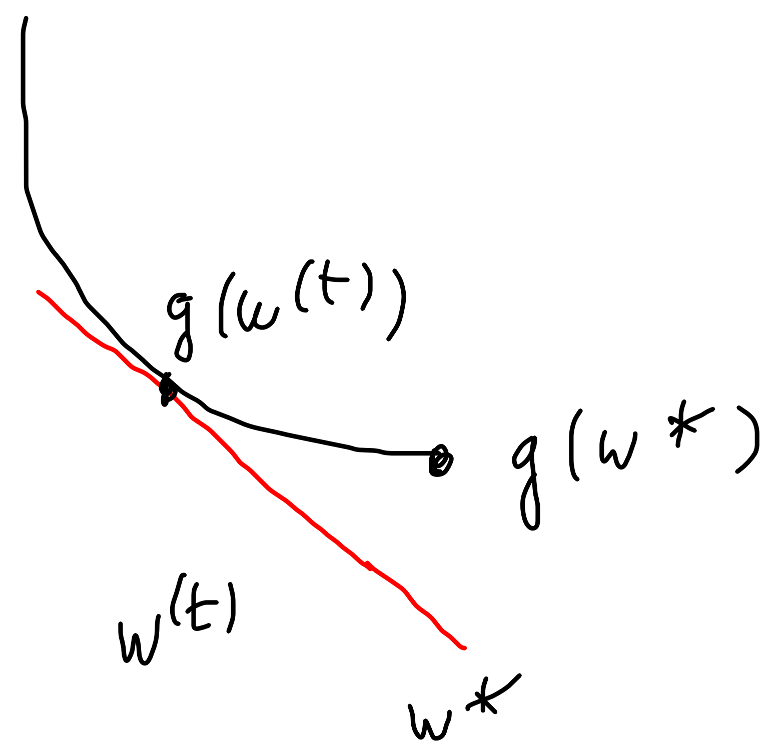
$$2\langle \eta \nabla g(w^{(t)}), w^{(t)} - w^* \rangle \leq \|w^{(t)} - w^*\|^2 - \|w^{(t+1)} - w^*\|^2 + \eta^2 G^2$$

$$\forall t = 1 \dots T$$

$$g(w^{GD}) - g(w^*) =$$

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

g è convessa $\rightarrow \leq \frac{1}{T} \sum_{t=1}^{T} \left[\nabla g(w^{(t)})^T (w^{(t)} - w^*) \right]$



$$g(w^*) \geq g(w^{(t)}) + \nabla g(w^{(t)})^T (w^* - w^{(t)})$$

$$\hookrightarrow g(w^{(t)}) - g(w^*) \leq + \nabla g(w^{(t)})^T (w^{(t)} - w^*)$$

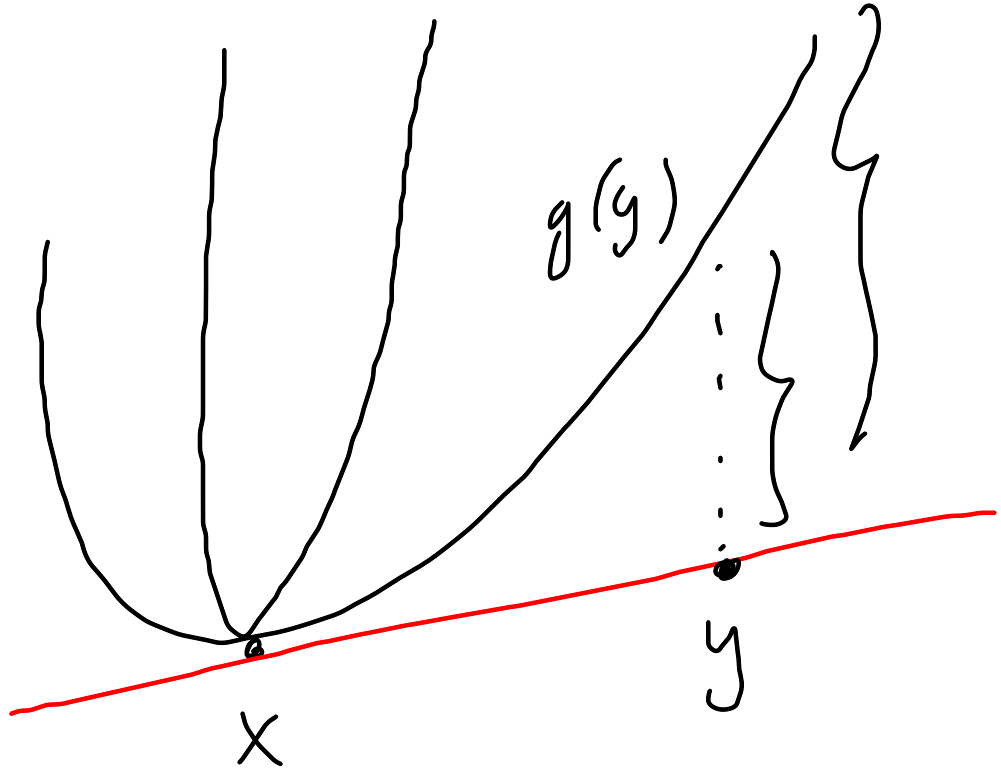
$$g(w^{GD}) - g(w^*) \le \frac{1}{T}\sum_{t=1}^{T} \frac{1}{2\eta}\left[\|w^{(t)} - w^*\|^2 - \|w^{(t+1)} - w^*\| + \eta^2 G^2 \right]$$

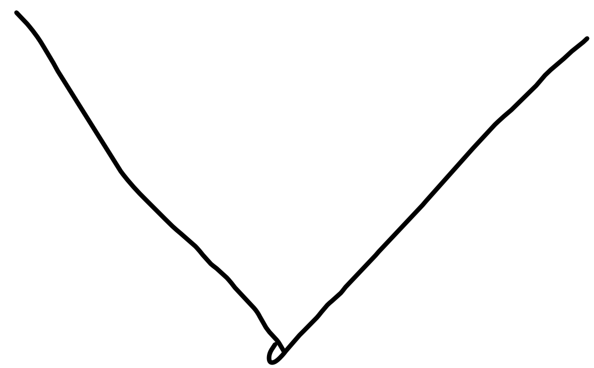
$$= \frac{1}{T}\sum_{t=1}^{T} \frac{1}{2\eta}\underbrace{\left[\|w^{(t)} - w^*\|^2 - \|w^{(t+1)} - w^*\|^2 \right]}_{\text{telescopice}} + \frac{\eta^{\cancel{2}} G^2}{2\cancel{\eta}}$$

$$= \frac{1}{T}\,\frac{1}{2\eta}\left(\overbrace{\|w^{(1)} - w^*\|^2}^{\le D^2} - \overbrace{\cancel{\|w^{(T+1)} - w^*\|^2}}^{\le 0} \right) + \frac{\eta G^2}{2}$$

$$\eta = \frac{D}{G\sqrt{T}}$$

$$= \frac{D^2}{2\eta T} + \frac{\eta G^2}{2} = \frac{D^{\cancel{2}} \cdot G \cancel{\sqrt{T}}}{2\cancel{D}\,\cancel{\sqrt{T}}} + \frac{D G^{\cancel{2}}}{\cancel{G}\sqrt{T} \cdot 2} = \frac{DG}{\sqrt{T}}$$


g(y)
y
x



$$f(z) = |z|$$

$$z \in \mathbb{R}$$

$$\frac{\partial f}{\partial z}(x) = \begin{cases} +1 & x > 0 \\ -1 & x < 0 \\ 0 & x = 0 \end{cases}$$

$$= \operatorname{sgn}(x)$$