5.3 Il percettrone

5.3.1 Terminazione dell'algoritmo del percettrone

Teorema 5.3.1. *Su un dataset linearmente separabile, l'algoritmo del percettrone termina dopo un numero finito di passi.*

Dimostrazione. Sia w^* un vettore di parametri che separa perfettamente gli esempi positivi e negativi, ovvero tale che $yw^{*\top}x > 0$ per ogni (x, y) nel dataset S.

L'idea della dimostrazione è di analizzare la quantità $\left\|w^{(t+1)} - \alpha w^*\right\|^2$ al variare di t (per un opportuno valore di $\alpha > 0$ precisato in seguito). Si consideri l'esempio $(x, y) \in S$ considerato al t-esimo passo dell'algoritmo. Se $yw^{(t)\top}x > 0$ (esempio classificato correttamente), il vettore dei parametri non viene modificato ($w^{(t+1)} = w^{(t)}$) e quindi

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

Se invece $yw^{(t)\top}x \leq 0$ (esempio misclassificato), $w^{(t+1)} = w^{(t)} + \eta yx$ per definizione dell'algoritmo. Il passo η in questo contesto ha solo l'effetto di scalare i vettori di esempio e non vi è perdita di generalità nel considerare $\eta = 1$. Si ottiene così

$$\begin{aligned}\left\|w^{(t+1)} - \alpha w^*\right\|^2 &= \left\|\left(w^{(t)} - \alpha w^*\right) + yx\right\|^2 \\ &= \left\|w^{(t)} - \alpha w^*\right\|^2 + \|yx\|^2 + 2\langle w^{(t)} - \alpha w^*, yx\rangle.\end{aligned}$$

Il prodotto scalare può essere maggiorato (usando $yw^{(t)\top}x \leq 0$)

$$2\langle w^{(t)} - \alpha w^*, yx\rangle = 2\langle w^{(t)}, yx\rangle - 2\alpha\langle w^*, yx\rangle \leq -2\alpha\langle w^*, yx\rangle,$$

ottenendo quindi

$$\left\|w^{(t+1)} - \alpha w^*\right\|^2 \leq \left\|w^{(t)} - \alpha w^*\right\|^2 + \|x\|^2 - 2\alpha\langle w^*, yx\rangle.$$

Si definiscano ora le costanti

$$\begin{aligned}\beta &\stackrel{\text{def}}{=} \max_{(x,y)\in S} \|x\| \\ \gamma &\stackrel{\text{def}}{=} \min_{(x,y)\in S} yw^{*\top}x \\ \alpha &\stackrel{\text{def}}{=} \beta^2/\gamma\end{aligned}$$

e si osservi che γ è positiva per definizione di w^*. Si può ora maggiorare

$$\begin{aligned}\left\|w^{(t+1)} - \alpha w^*\right\|^2 &\leq \left\|w^{(t)} - \alpha w^*\right\|^2 + \|x\|^2 - 2\alpha\langle w^*, yx\rangle \\ &\leq \left\|w^{(t)} - \alpha w^*\right\|^2 + \beta^2 - 2\alpha\gamma \\ &= \left\|w^{(t)} - \alpha w^*\right\|^2 - \beta^2.\end{aligned}$$

La quantità $\left\| w^{(t)} - \alpha w^* \right\|^2$ diminuisce quindi di almeno β^2 ad ogni misclassificazione, ma non potendo essa diventare negativa, il numero di iterazioni in cui si ha una misclassificazione deve essere finito. Tale numero può essere al più

$$\frac{\left\| w^{(1)} - \alpha w^* \right\|^2}{\beta^2}.$$

Quando $w^{(1)} = 0$, tale numero è al più $\alpha^2 \left\| w^* \right\|^2 / \beta^2 = \beta^2 \left\| w^* \right\|^2 / \gamma^2$. □