Neurone artificiale basato sulla regressione logistica

Ridefiniamo i target $y^i = 0, 1$

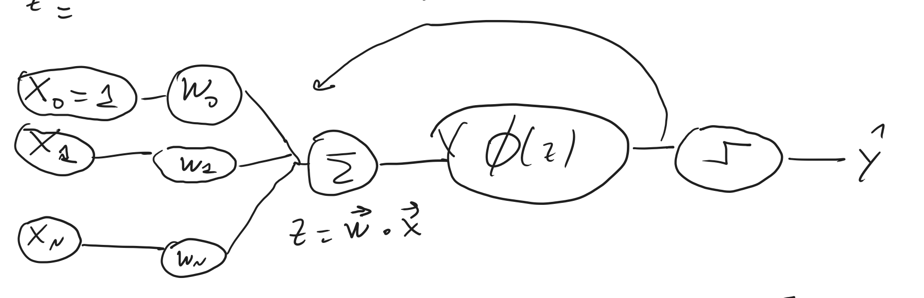
e definiamo la funzione scalino

$$6(\phi(z)) = \begin{cases} 1 & \text{se } \phi(z) > \frac{1}{2} \\ 0 & \text{se } \phi(z) < \frac{1}{2} \end{cases}$$

Ricordiamo che

$z =$

funzione d'errore



$$l(p) = \log\left(\frac{p}{1-p}\right) \qquad p \in [0, 1]$$

$$l(\phi(z)) = z$$

$$\log\left(\frac{\phi(z)}{1-\phi(z)}\right) = z$$

$$\frac{\phi(z)}{1-\phi(z)} = e^z$$

$$\phi(z) = e^z(1-\phi(z))$$

$$\phi(z)(1+e^z) = e^z$$

$$\phi(z) = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}} \qquad \text{funzione sigmoide}$$

Definiamo una funzione PERFORMANCE

$$L(\vec{w}) = \prod_{i=1, N_{sample}} (\phi(z^i))^{y^i} (1-\phi(z^i))^{(1-y^i)}$$

$L(\vec{w})$ andrebbe massimizzata, ma si preferisce

$$J(\vec{w}) = -\log L(\vec{w})$$

$$J(\vec{w}) = \sum_{i=1, N_{sample}} -y^i \log(\phi(z^i)) - (1-y^i)\log(1-\phi(z^i))$$

$$\frac{\partial J(\vec{w})}{\partial w_j} = \sum_{i=1, N_{samples}} -y^i \frac{1}{\phi(z^i)} \frac{\partial \phi(z^i)}{\partial w_j} - (1-y^i)\frac{1}{(1-\phi(z^i))}\frac{\partial \phi(z^i)}{\partial w_j}$$

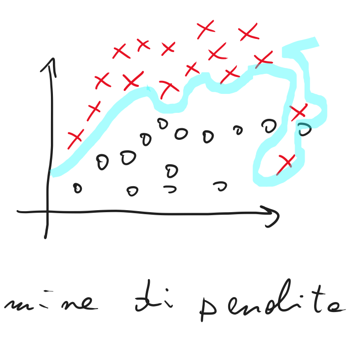
Dove

$$\frac{\partial \phi(z^i)}{\partial w_j} = \frac{\partial}{\partial w_j}\left(\frac{1}{1+e^{-z^i}}\right) = -\frac{1}{(1+e^{-z^i})^2}(-e^{-z^i}) = \frac{x_j^i e^{-z^i}}{(1+e^{-z^i})^2}$$

$$= \phi(z^i)(1-\phi(z^i))\, x_j^i$$

$$\frac{\partial J}{\partial w_j} = \sum_{i=1, N_{samples}} (-y^i + \phi(z^i))\, x_j^i$$

NOTA: vogliamo evitare l'over-fitting

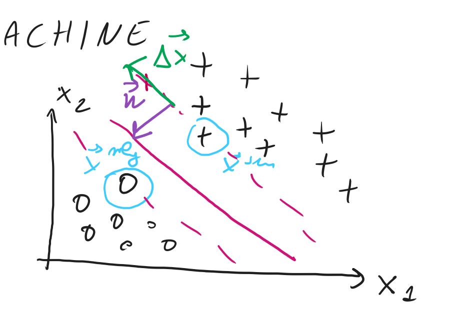


Si introduce un termine di penalità nella funzione d'errore

$$J'(\vec{w}) = C\, J(\vec{w}) + \frac{1}{2}\vec{w}\cdot\vec{w}$$

C: costante positiva

SUPPORT VECTOR MACHINE



Notazione

$\vec{x} = (x_1, \ldots, x_N)$

$\vec{w} = (w_0, \ldots, w_N)$

$y = -1, 1$

Supponiamo il problema sia linearmente separabile

L'iperpiano che separa le soluzioni è dato da un'equazione del tipo

$$\phi(\vec{x}) = 0 \quad \text{con } \phi(\vec{x}) = \vec{w}\cdot\vec{x} + w_0$$

I vettori più vicini al piano di separazione sono detti support vector, li chiamo $\vec{x}^{sup}$ e $\vec{x}^{neg}$

Voglio che $\phi(\vec{x}^{sup}) = 1$ e $\phi(\vec{x}^{neg}) = -1$

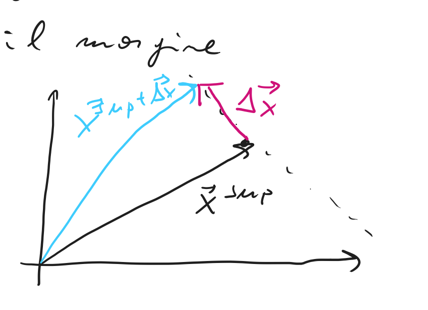
$$\begin{cases} \vec{w}\cdot\vec{x}^{sup} + w_0 = 1 \\ \vec{w}\cdot\vec{x}^{neg} + w_0 = -1 \end{cases}$$

Consideriamo i piani $\phi(\vec{x}) = 1$ e $\phi(\vec{x}) = -1$ detti margini

Mostriamo che $\vec{w}$ è $\perp$ ai margini

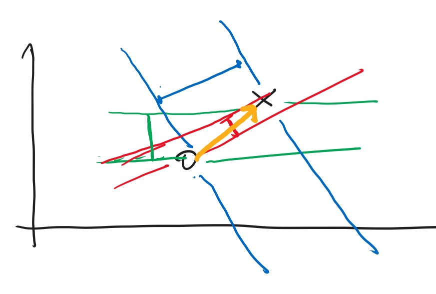
Sia $\Delta\vec{x}$ vettore entro il margine

$\vec{x}^{sup} + \Delta\vec{x}$ individua un punto nel margine



$$\phi(\vec{x}^{sup} + \Delta\vec{x}) = 1$$

$$\underbrace{\vec{w}\cdot\vec{x}^{sup} + w_0}_{=1} + \vec{w}\cdot\Delta\vec{x} = 1 \quad \Rightarrow \quad \vec{w}\cdot\vec{\Delta x} = 0$$



Vogliamo scegliere $\vec{w}$ in maniera da massimizzare

d distanza tra i margini:

$$d = (\vec{x}^{sup} - \vec{x}^{neg}) \cdot \frac{\vec{w}}{|\vec{w}|}$$

$$= \frac{\vec{x}^{sup}\cdot\vec{w}}{|\vec{w}|} - \frac{\vec{x}^{neg}\cdot\vec{w}}{|\vec{w}|} \pm \frac{w_0}{|\vec{w}|}$$

$$= \frac{\vec{x}^{sup}\cdot\vec{w} + w_0}{|\vec{w}|} - \frac{(\vec{x}^{neg}\cdot\vec{w} + w_0)}{|\vec{w}|} = \frac{2}{|\vec{w}|}$$

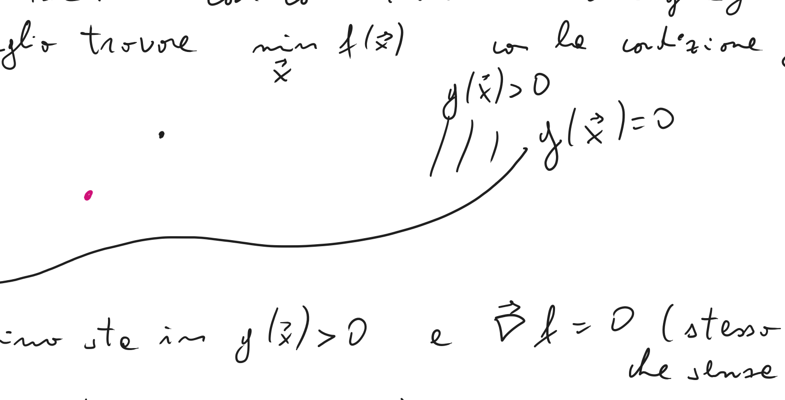
Questo è equivalente a minimizzare

$$J(\vec{w}) = \frac{1}{2}\vec{w}\cdot\vec{w}$$

che va minimizzata con la condizione $y^i\underbrace{(w_0 + \vec{w}\cdot x^i)}_{\phi(\vec{x}^i)} \geq 1$

Minimizzazione con condizioni di disuguaglianza

Voglio trovare $\min_{\vec{x}} f(\vec{x})$ con la condizione $g(\vec{x}) \geq 0$



il minimo sta in $g(\vec{x}) > 0$ e $\vec{\nabla} f = 0$ (stesso minimo che senza condizione)

altrimenti deve stare su $g(\vec{x}) = 0$

$$p^* = \min_{\vec{x}} f(\vec{x}) \qquad \text{con } g(\vec{x}) \geq 0$$

$$J(\vec{x}) = \begin{cases} f(\vec{x}) & \text{se } g(\vec{x}) \geq 0 \\ +\infty & \text{altrimenti} \end{cases}$$

$$p^* = \min_{\vec{x}} J(\vec{x})$$

$$H(\vec{x}, \lambda) = f(\vec{x}) - \lambda\, g(\vec{x}) \qquad \lambda \geq 0$$

$$\max_{\substack{\lambda \\ \lambda \geq 0}} H(\vec{x}, \lambda) = J(\vec{x})$$

$$p^* = \min_{\vec{x}} \max_{\substack{\lambda \\ \lambda \geq 0}} H(\vec{x}, \lambda)$$

$$H(\vec{x}, \lambda) \leq J(\vec{x})$$

$$\min_{\vec{x}} H(\vec{x}, \lambda) \leq \min_{\vec{x}} J(\vec{x})$$

$$\max_{\lambda} \min_{\vec{x}} H(\vec{x}, \lambda) \leq p^*$$

Supponiamo sia scrivibile con '=' strong duality

Applicazione a SVM

$$H = \frac{1}{2}\vec{w}\cdot\vec{w} - \sum_{i=1, N_{samples}} \lambda_i y^i (\vec{w}\cdot x^i + w_0)$$

Minimizziamo prima rispetto a $\vec{w}$ e w_0

$$\frac{\partial H}{\partial \vec{w}} = \vec{w} - \sum_{i=1, N_{samples}} \lambda_i y^i \vec{x}^i = 0 \quad \Rightarrow \vec{w} = \sum_i \lambda_i y^i x^i$$

$$\frac{\partial H}{\partial w_0} = \sum_{i=1, N_{samples}} \lambda_i y^i = 0$$

$$H = \frac{1}{2}\left(\sum_i \lambda_i y^i \vec{x}^i\right)\cdot\left(\sum_j \lambda_j y^j \vec{x}^j\right) - \sum_i y^i \lambda_i \vec{x}^i \cdot \left(\sum_j y^j \lambda_j \vec{x}^j\right)$$

$$= -\frac{1}{2}\sum_{i,j=1, N_{samples}} \lambda_i \lambda_j y^i y^j \vec{x}^i \cdot \vec{x}^j$$

che va MASSIMIZZATA ri

Problemi non line