Sigmoide.

Un neurone artificiale ha una funzione di attivazione descritta da una curva sigmoide di equazione

$$\sigma(x) = \frac{1}{2} + \frac{\arctan(x)}{\pi}.$$

Si vuole approssimare la sigmoide con una funzione lineare a tratti costituita da tre segmenti rettilinei.

Si vuole minimizzare la massima differenza in valore assoluto tra il valore della sigmoide e quello della funzione lineare a tratti.

Formulare il problema, classificarlo e risolverlo.

Discutere l'ottimalità e l'unicità della soluzione ottenuta.

Suggerimento: la derivata prima della funzione $f(x) = \arctan(x)$ è la funzione $f'(x) = \dfrac{1}{1+x^2}$.

Soluzione commentata.

Il problema non richiede alcun dato oltre all'equazione della sigmoide.

Si noti anzitutto che la sigmoide è compresa tra 0 e 1 ed è simmetrica rispetto al punto $(0, \dfrac{1}{2})$. Quindi anche la funzione lineare a tratti che la appprossima in modo ottimale ha la stessa proprietà e passa per il centro di simmetria.

Per evitare che la differenza tra la sigmoide e la funzione lineare a tratti possa crescere all'infinito, quando l'input x del neurone tende a $+\infty$ o $-\infty$, è necessario che il primo ed il terzo segmento della funzione cercata siano orizzontali. Sia δ_1 la loro distanza dai valori estremi 0 e 1.

Il secondo tratto della funzione cercata è caratterizzato dal suo coefficiente angolare m.

Data la simmetria, possiamo considerare solo il semispazio $x \geq 0$.

Dai valori di δ_1 e di m si possono facilmente ricavare le cooordinate del punto di discontinuità D:

$$x_D = \frac{\frac{1}{2} - \delta_1}{m} \qquad y_D = 1 - \delta_1.$$

A destra di D, per x che tende a ∞ la sigmoide tende a 1 e quindi la differenza tra la sigmoide e la funzione lineare a tratti non supera il valore limite δ_1.

In D la funzione lineare a tratti è sopra la sigmoide. Sia δ_2 lo scostamento tra le due funzioni in x_D:

$$\delta_2 = y_D - \sigma(x_D).$$

A sinistra di D, esiste un punto T in cui la sigmoide è sopra alla funzione cercata e la differenza tra le due è massima. Tale punto è quello in cui la tangente alla sigmoide ha coefficiente angolare uguale a m.

$$\sigma'(x_T) = m.$$

Sia δ_3 lo scostamento tra le due funzioni in x_T:

$$\delta_3 = \sigma(x_T) - \left(\frac{1}{2} + m x_T\right)$$

Trattandosi di un problema con funzione obiettivo min-max (minimizzare il massimo scostamento tra le funzioni), l'ottimo si ha quando gli scostamenti sono uguali, cioè quando $\delta_1 = \delta_2 = \delta_3$.

Si può quindi formulare il problema come problema di ottimizzazione come

$$\text{minimize } z = \Delta$$

con vincoli

$$\Delta \geq \delta_1 \quad \Delta \geq \delta_2 \quad \Delta \geq \delta_3$$

oppure come problema di esistenza, imponendo direttamente $\delta_1 = \delta_2 = \delta_3$ senza obiettivo.

Il modello del problema è comunque di programmazione non-lineare ed è convesso. La soluzione calcolata dal solutore è ottima ed è unica.

I valori ottimi sono $m^* = 0.187631$ e $\delta^* = 0.0647829$. I punti T e D hanno ascisse $x_T = 0.834545$ e $x_D = 2.31953$.