Desarrollo del algoritmo de aprendizaje Backpropagation para un Perceptrón Multicapa (MLP) con 1 capa oculta

Ajuste de pesos sinápticos según el descenso por el gradiente:

$$\delta w'_{kj} = -\varepsilon \frac{\partial E}{\partial w'_{kj}} \qquad (1)$$

$$\delta w_{ji} = -\varepsilon \frac{\partial E}{\partial w_{ji}} \qquad (2)$$

Salida de la red MLP:

$$z_k^\mu = g\left(\sum_j w'_{kj} \underbrace{f\left(\sum_i w_{ji} x_i^\mu - \theta_j\right)}_{y_j^\mu} - \theta'_k\right) \qquad (3)$$

Función de error (error cuadrático medio):

$$E(w_{ji}, w'_{kj}) = \frac{1}{2} \sum_\mu \sum_k \left[t_k^\mu - z_k^\mu\right]^2 \qquad (4)$$

Reemplazando (3) en (4)

$$E(w_{ji}, w'_{kj}) = \frac{1}{2} \sum_\mu \sum_k \left[t_k^\mu - g\left(\underbrace{\sum_j w'_{kj} y_j^\mu - \theta'_k}_{h'^\mu_k}\right)\right]^2 = \frac{1}{2} \sum_\mu \sum_k \left[\underbrace{t_k^\mu - g\left(h'^\mu_k\right)}_{h''^\mu_k}\right]^2 \qquad (5)$$

Cálculo de (1): Derivando (5) respecto a cada peso w'_{kj} y aplicando la regla de la cadena, tenemos:

$$\frac{\partial E}{\partial w'_{kj}} = \frac{\partial E}{\partial h''^\mu_k} \frac{\partial h''^\mu_k}{\partial h'^\mu_k} \frac{\partial h'^\mu_k}{\partial w'_{kj}} = \frac{1}{2} 2 \sum_\mu \left[t_k - g\left(h'^\mu_k\right)\right] \left(-\frac{\partial g\left(h'^\mu_k\right)}{\partial h'^\mu_k}\right) \underbrace{\frac{\partial h'^\mu_k}{\partial w'_{kj}}}_{y_j^\mu \text{ (por (5))}} =$$

$$= -\sum_\mu \left[t_k - g\left(h'^\mu_k\right)\right] \frac{\partial g\left(h'^\mu_k\right)}{\partial h'^\mu_k} y_j^\mu \qquad (6)$$

Reemplazando (6) en (1) tenemos

$$\delta w'_{kj} = -\varepsilon \frac{\partial E}{\partial w'_{kj}} = \varepsilon \sum_\mu \underbrace{\left[t_k - g\left(h'^\mu_k\right)\right] \frac{\partial g\left(h'^\mu_k\right)}{\partial h'^\mu_k}}_{\Delta'^\mu_k} y_j^\mu$$

Por lo tanto

$$\delta w'_{kj}=\varepsilon\sum_{\mu}\Delta'^{\mu}_{k}\,y^{\mu}_{j}\quad\left|\begin{array}{l}\Delta'^{\mu}_{k}=\left[t^{\mu}_{k}-g\left(h'^{\mu}_{k}\right)\right]\dfrac{\partial g\left(h'^{\mu}_{k}\right)}{\partial h'^{\mu}_{k}}\\[2ex] h'^{\mu}_{k}=\displaystyle\sum_{j}w'_{kj}\,y^{\mu}_{j}-\theta'_{k}\end{array}\right.$$

De manera similar se calcula (2). Expandiendo y^{μ}_{j} en (5) tenemos:

$$E\left(w_{ji},w'_{kj}\right)=\frac{1}{2}\sum_{\mu}\sum_{k}\left[\underbrace{t^{\mu}_{k}-g\left(\underbrace{\sum_{j}w'_{kj}\,f\left(\underbrace{\sum_{i}w_{ji}\,x^{\mu}_{i}-\theta_{j}}_{h^{\mu}_{j}}\right)-\theta'_{k}}_{h'^{\mu}_{k}}\right)}_{h''^{\mu}_{k}}\right]^{2}\qquad(7)$$

Cálculo de (2): derivando (5) respecto a cada peso w_{ji} y aplicando la regla de la cadena, tenemos:

$$\frac{\partial E}{\partial w_{ji}}=\frac{\partial E}{\partial h''^{\mu}_{k}}\;\frac{\partial h''^{\mu}_{k}}{\partial h'^{\mu}_{k}}\;\frac{\partial h'^{\mu}_{k}}{\partial h^{\mu}_{j}}\;\frac{\partial h^{\mu}_{j}}{\partial w_{ji}}=\frac{1}{2}2\sum_{\mu}\sum_{k}\left[\left[t_{k}-g\left(h'^{\mu}_{k}\right)\right]\left(-\frac{\partial g\left(h'^{\mu}_{k}\right)}{\partial h'^{\mu}_{k}}\right)\;w'_{kj}\;\frac{\partial f\left(h^{\mu}_{j}\right)}{\partial h^{\mu}_{j}}\;\underbrace{\frac{\partial h^{\mu}_{j}}{\partial w_{ji}}}_{x^{\mu}_{i}\,(por(7))}\right]=$$

$$=-\sum_{\mu}\sum_{k}\left[\left[t_{k}-g\left(h'^{\mu}_{k}\right)\right]\frac{\partial g\left(h'^{\mu}_{k}\right)}{\partial h'^{\mu}_{k}}\;w'_{kj}\;\frac{\partial f\left(h^{\mu}_{j}\right)}{\partial h^{\mu}_{j}}\right]\;x^{\mu}_{i}\qquad(8)$$

Reemplazando (8) en (2) tenemos

$$\delta w_{ji}=-\varepsilon\frac{\partial E}{\partial w_{ji}}=\varepsilon\sum_{\mu}\;\underbrace{\sum_{k}\;\underbrace{\left[t_{k}-g\left(h'^{\mu}_{k}\right)\right]\;\frac{\partial g\left(h'^{\mu}_{k}\right)}{\partial h'^{\mu}_{k}}}_{\Delta'^{\mu}_{k}}\;w'_{kj}\;\frac{\partial f\left(h^{\mu}_{j}\right)}{\partial h^{\mu}_{j}}}_{\Delta^{\mu}_{j}}\;x^{\mu}_{i}$$

Por lo tanto

$$\delta w_{ji}=\varepsilon\sum_{\mu}\Delta^{\mu}_{j}\,x^{\mu}_{i}\quad\left|\begin{array}{l}\Delta^{\mu}_{j}=\left[\displaystyle\sum_{k}\Delta'^{\mu}_{k}\,w'_{kj}\right]\;\dfrac{\partial f\left(h^{\mu}_{j}\right)}{\partial h^{\mu}_{j}}\\[2ex]\Delta'^{\mu}_{k}=\left[t^{\mu}_{k}-g\left(h'^{\mu}_{k}\right)\right]\dfrac{\partial g\left(h'^{\mu}_{k}\right)}{\partial h'^{\mu}_{k}}\\[2ex] h'^{\mu}_{k}=\displaystyle\sum_{j}w'_{kj}\,y^{\mu}_{j}-\theta'_{k}\\[2ex] h^{\mu}_{j}=\displaystyle\sum_{i}w_{ji}\,x^{\mu}_{i}-\theta_{j}\end{array}\right.$$