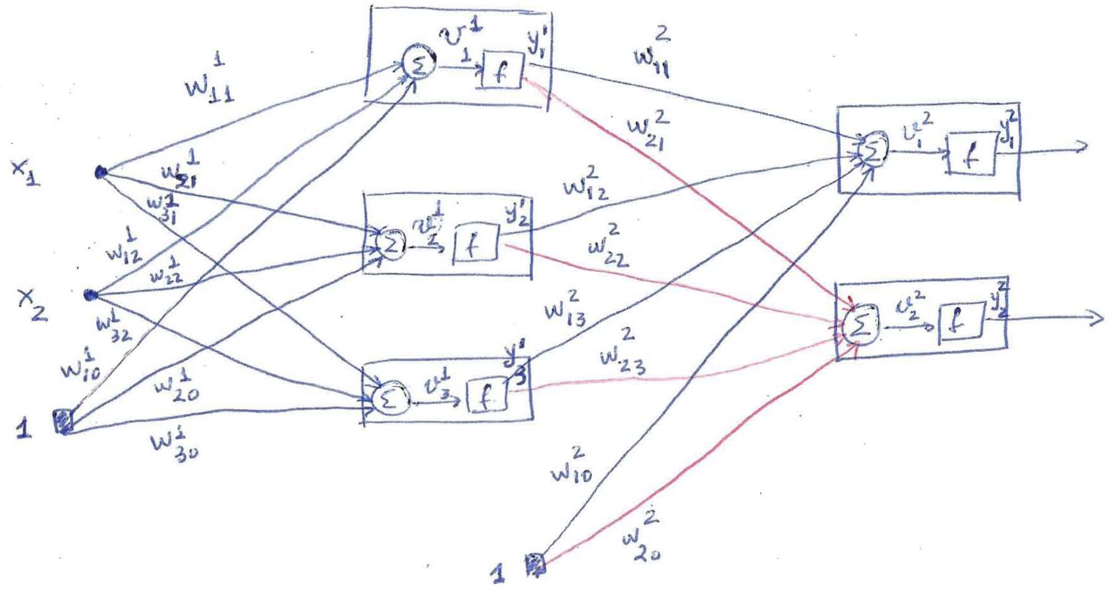
1

\# of layers : $L=2$

\# of 1st layer nodes : $K_1=3$

\# of 2nd layer nodes : $K_2=2$.



$$J = \sum_{i=1}^{N} \mathcal{E}(i) \quad (1), \qquad \mathcal{E}(i) = \frac{1}{2}\sum_{m=1}^{K_2} e_m^2(i) = \frac{1}{2}\sum_{m=1}^{K_2}\left(y_m(i) - \hat{y}_m(i)\right)^2. \quad (2)$$

$$\underline{w}_j^r = \left[w_{j0}^r, w_{j1}^r, \dots, w_{jK_{r-1}}^r \right]^T, \qquad y^{r-1}(i) = \left[1, y_1^{r-1}(i), \dots, y_{K_{r-1}}^{r-1}(i) \right]^T.$$

$$\underline{w}_j^r(\text{new}) = \underline{w}_j^r(\text{old}) + \Delta \underline{w}_j^r \quad (3) \qquad \Delta \underline{w}_j^r = -\mu \frac{\partial J}{\partial \underline{w}_j^r} \quad (4)$$

$$v_j^r(i) = \sum_{k=1}^{K_{r-1}} w_{jk}^r\, y_k^{r-1}(i) = \underline{w}_j^{r\,T} \cdot \underline{y}^{r-1}(i) \quad (5) \qquad y_j^r(i) = f\left(v_j^r(i)\right) \quad (5a)$$

$$\left.\begin{aligned} \frac{\partial \mathcal{E}(i)}{\partial w_j^r} &= \frac{\partial \mathcal{E}(i)}{\partial v_j^r(i)} \cdot \frac{\partial v_j^r(i)}{\partial \underline{w}_j^r} \\ \frac{\partial v_j^r(i)}{\partial \underline{w}_j^r} &= \underline{y}^{r-1}(i) \end{aligned}\right\} \longrightarrow \frac{\partial \mathcal{E}(i)}{\partial w_j^r} = \frac{\partial \mathcal{E}(i)}{\partial v_j^r(i)} \cdot y^{r-1}(i) \quad (6)$$

$$(4) + (6) \longrightarrow \Delta \underline{w}_j^r = -\mu \sum_{i=1}^{N} \frac{\partial \mathcal{E}(i)}{\partial v_j^r(i)}\, y^{r-1}(i) \equiv -\mu \sum_{i=1}^{N} \delta_j^r(i) \cdot y^{r-1}(i) \quad (7)$$

$$\delta_j^r(i) \equiv \frac{\partial \mathcal{E}(i)}{\partial v_j^r(i)} \quad (8)$$

Υπολογισμός των $\delta_j^r(i)$

1. $r = L$

$$\delta_j^L(i) = \frac{\partial \mathcal{E}(i)}{\partial v_j^L(i)} = \frac{\partial}{\partial v_j^L(i)}\left[\frac{1}{2}\sum_{m=1}^{k_L} e_m^2(i)\right] = \frac{\partial}{\partial v_j^L(i)}\left[\frac{1}{2}\sum_{m=1}^{k_L}\left(f(v_m^L(i)) - y_m(i)\right)^2\right] =$$

$$= \left(f(v_j^L(i)) - y_j(i)\right) \cdot f'(v_j^L(i)) = e_j(i) \cdot f'(v_j^L(i)).$$

$$\boxed{\delta_j^L(i) = e_j(i) \cdot f'(v_j^L(i))} \quad (9) \quad , \quad e_j(i) = \left(f(v_m^L(i)) - y_m(i)\right)$$

2. $r < L$

$$\delta_j^{r-1}(i) = \frac{\partial \mathcal{E}(i)}{\partial v_j^{r-1}(i)} = \sum_{k=1}^{k_r} \frac{\partial \mathcal{E}(i)}{\partial v_k^r(i)} \cdot \frac{\partial v_k^r(i)}{\partial v_j^{r-1}(i)}$$

$$\Rightarrow \delta_j^{r-1}(i) = \sum_{k=1}^{k_r} \delta_k^r(i) \frac{\partial v_k^r(i)}{\partial v_j^{r-1}(i)} \quad (10)$$

$$\frac{\partial v_k^r(i)}{\partial v_j^{r-1}(i)} = \frac{\partial}{\partial v_j^{r-1}(i)}\left[\sum_{m=0}^{k_{r-1}} w_{km}^r \, y_m^{r-1}(i)\right] = w_{kj}^r \cdot \frac{\partial f(v_j^{r-1}(i))}{\partial v_j^{r-1}(i)} =$$

$$= w_{kj}^r \, f'(v_j^{r-1}(i)) \quad (11)$$

$$(10)+(11) \rightarrow \boxed{\delta_j^{r-1}(i) = \left[\sum_{k=1}^{k_r} \delta_k^r(i) \cdot w_{kj}^r\right] \cdot f'(v_j^{r-1}(i))} \quad (12)$$

$$\boxed{\delta_j^{r-1}(i) = e_j^{r-1}(i) \cdot f'(v_j^{r-1}(i))} \quad (13) \qquad e_j^{r-1}(i) = \sum_{k=1}^{k_r} \delta_k^r(i) \cdot w_{kj}^r \quad (13\alpha)$$

Ο αλγόριθμος

1. Αρχικοποίησε όλα τα βάρη σε μικρές τυχαίες τιμές
2. Εμπρόσθιοι υπολογισμοί: Για κάθε διάνυσμα δεδομένων $x(i)$, $i=1,\dots,N$ υπολόγισε τα (a) $v_j^r(i)$, $y_j^r(i) = f(v_j^r(i))$, $j=1,\dots,k_r$, $r=1,\dots,L$. από τις (5), (5α)
   (b) $e_j(i)$, $j=1,\dots,k_L$
2. Οπίσθιοι υπολογισμοί: Για κάθε $i=1,\dots,N$ υπολόγισε τα:
   (α) $\delta_j^L(i)$, $j=1,\dots,k_L$ από την (9)
   (β) $\delta_j^{r-1}(i)$, $j=1,\dots,k_r$, $r=L, L-1,\dots,2$ από την (12) ή τις (13), (13α)
3. Ενημέρωση βαρών: για $r=1,\dots,L$, $j=1,\dots,k_r$ με χρήση των (3), (7)