数理科学特論Ａ　第１０回付録
誤差逆伝播法が最急降下になっていることの証明

　ある入出力例における，所望の出力例と現実の出力（第n層でのニューロンの状態）の誤差の２乗和を

$$E = \frac{1}{2}\sum_{l}(x_l^n - y_l)^2 \tag{A1}$$

と定義します．このとき，Eは各結合重みの関数です．いま，結合重み$w_{jl}^{k-1\,k}$ 方向のEの微分を求めると

$$\frac{\partial E}{\partial w_{jl}^{k-1\,k}} = \frac{\partial E}{\partial S_l^k}\frac{\partial S_l^k}{\partial w_{jl}^{k-1\,k}} \tag{A2}$$

となり，また(7)式から

$$\begin{aligned}\frac{\partial S_l^k}{\partial w_{jl}^{k-1\,k}} &= \frac{\partial}{\partial w_{jl}^{k-1\,k}}\sum_{i} w_{il}^{k-1\,k}x_i^{k-1}\\ &= x_j^{k-1}\end{aligned} \tag{A3}$$

となります．そこで

$$\delta_l^k = \frac{\partial E}{\partial S_l^k} \tag{A4}$$

とおいて, (A3)(A4)式を(A2)式に代入すると

$$\frac{\partial E}{\partial w_{jl}^{k-1\,k}} = \delta_l^k x_j^{k-1} \tag{A5}$$

となります．この式を結合重みを修正する方法を表す式である

$$w'^{k-1\,k}_{jl} = w_{jl}^{k-1\,k} - \varepsilon\delta_l^k x_j^{k-1} \tag{13}$$

と比較すると, (13)式は結合重みを$\partial E / \partial w_{jl}^{k-1\,k}$ 方向，すなわち$-\,\mathrm{grad}\,E$の方向に変化させており, (13)式による結合重みの修正はEの最急降下になっていることがわかります.

　さて, (A4)式のδは，「考え方」[1][2]から導いた本文のδと同じものであることを示しましょう．出力層以外，すなわち$k \neq n$のとき，

$$\frac{\partial E}{\partial S_j^k} = \sum_{l}\frac{\partial E}{\partial S_l^{k+1}}\cdot\frac{\partial S_l^{k+1}}{\partial x_j^k}\cdot\frac{dx_j^k}{dS_j^k} \tag{A6}$$

と変形します．本文の(10)式から

$$\frac{dx_j^k}{dS_j^k} = f'(S_j^k) \tag{10}$$

です．また(7)式の添え字を付け替えると

$$S_l^{k+1} = \sum_{i} w_{il}^{k\,k+1}x_i^k \tag{A8}$$

ですから

$$\frac{\partial S_l^{k+1}}{\partial x_j^k} = w_{jl}^{k\,k+1} \tag{A9}$$

となります. (10)式, (A9)式を(A6)式に代入すると

$$\frac{\partial E}{\partial S_j^k} = \sum_l \frac{\partial E}{\partial S_l^{k+1}} \cdot w_{jl}^{k\,k+1} \cdot f'(S_j^k) \tag{A10}$$

となり, (A4)式の δ の定義から

$$\delta_j^k = \frac{\partial E}{\partial S_j^k},\ \delta_l^{k+1} = \frac{\partial E}{\partial S_l^{k+1}} \tag{A11}$$

ですから，これを(A10)式に代入すると

$$\begin{aligned} \delta_j^k &= \sum_l \delta_l^{k+1} \cdot w_{jl}^{k\,k+1} \cdot f'(S_j^k) \\ &= \left[\sum_l (w_{jl}^{k\,k+1}\delta_l^{k+1})\right] f'(S_j^k) \end{aligned} \tag{A12}$$

となり本文の δ の定義と同じになります.

出力層，すなわち $k = n$ のときは,

$$\frac{\partial E}{\partial S_l^n} = \frac{\partial E}{\partial x_l^n} \cdot \frac{dx_l^n}{dS_l^n} \tag{A13}$$

と変形します．誤差２乗和の定義である(A1)式から

$$\begin{aligned} \frac{\partial E}{\partial x_l^n} &= \frac{\partial}{\partial x_l^n}\left\{\frac{1}{2}\sum_l (x_l^n - y_l)^2\right\} \\ &= x_l^n - y_l \end{aligned} \tag{A14}$$

となります．これを(A13)式に代入すると

$$\frac{\partial E}{\partial S_l^n} = (x_l^n - y_l) \cdot \frac{dx_l^n}{dS_l^n} \tag{A15}$$

となり, (A4)式の δ の定義から

$$\delta_l^n = \frac{\partial E}{\partial S_l^n} \tag{A16}$$

で，また(10)式から

$$\frac{dx_l^n}{dS_l^n} = f'(S_l^n) \tag{A17}$$

ですから, (A16)式, (A17)式を(A15)式に代入すると

$$\delta_l^n = (x_l^n - y_l) f'(S_l^n) \tag{A18}$$

となり本文の(12)式の δ の定義と同じになります.