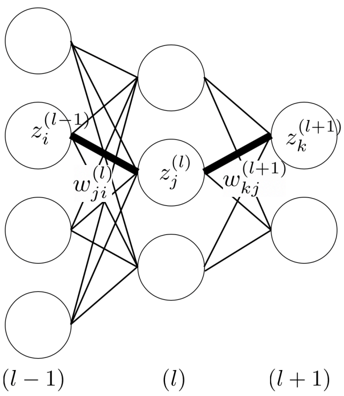
二日目課題: バックプロパゲーション（筆記課題）

ニューラルネットワークの第 $l-1$ 層第 i ユニットの出力を $z_i^{(l-1)}$ で表し、同ユニットから第 l 層第 j ユニットへの結線の重みを $w_{ji}^{(l)}$ で表す。第 l 層第 j ユニットの出力は $z_j^{(l)} = f(u_j^{(l)})$ であり、$u_j^{(l)} = \sum_i w_{ji}^{(l)} z_i^{(l-1)}$ である。同様に、第 $l+1$ 層第 k ユニットの出力は $z_k^{(l+1)} = f(u_k^{(l+1)})$ であり、$u_k^{(l+1)} = \sum_j w_{kj}^{(l+1)} z_j^{(l)}$ である。ここで $f(u)$ は活性化関数を表す。ネットワークが最小化するコスト関数を E_n で表す。コスト最小化のために次式の右辺を求めたい。

$$\frac{\partial E_n}{\partial w_{ji}^{(l)}} = \frac{\partial E_n}{\partial u_j^{(l)}} \frac{\partial u_j^{(l)}}{\partial w_{ji}^{(l)}} \tag{1}$$

以下の問いに答えよ。



内積の係数(重み)による微分

(a) $\underline{\partial u_j^{(l)}/\partial w_{ji}^{(l)}}$ をしかるべき z を用いて表せ。

$z_i^{(l-1)}$

(b) $f(u)$ の微分を $\underline{f'(u)}$ で表す。$\partial u_k^{(l+1)}/\partial u_j^{(l)}$ をしかるべき w と $f'()$ を用いて表せ。

$$\frac{\partial u_k^{(l+1)}}{\partial z_j^{(l)}} \frac{\partial z_j^{(l)}}{\partial u_j^{(l)}} = w_{kj}^{(l+1)} f'(u_j^{(l)})$$

(c) 次式が成立する

$$\frac{\partial E_n}{\partial u_j^{(l)}} = \sum_k \frac{\partial E_n}{\partial u_k^{(l+1)}} \frac{\partial u_k^{(l+1)}}{\partial u_j^{(l)}}$$

$\delta_j^{(l)} = \partial E_n / \partial u_j^{(l)}$ で表すと、次式が成立する。

$$\delta_j^{(l)} = \sum_k \delta_k^{(l+1)} \left(\frac{\partial u_k^{(l+1)}}{\partial u_j^{(l)}} \right)$$

右辺の第二項（カッコの中）は上記 2 の答えで計算できるため、$l+1$ 層の $\delta_k^{(l+1)}$ から l 層の $\delta_j^{(l)}$ を計算できる。式 (1) は次式のように書き直すことが出来る：

$$\frac{\partial E_n}{\partial w_{ji}^{(l)}} = \delta_j^{(l)} \left(\frac{\partial u_j^{(l)}}{\partial w_{ji}^{(l)}} \right)$$

上式の右辺第二項（カッコの中）は上記 1 の答えで計算できる。これらにより E_n の微分を計算できる。このように $\delta_j^{(l)}$ を出力層から入力層に向かって求めていく計算方法を誤差逆伝播法 (back propagation) という。右下図に示す 3 層のニューラルネットワークで回帰の問題を解く。3 層目は 2 つのニューロン ($k = 1, 2$) で構成されており、活性化関数は恒等関数、すなわち、$z_k^{(3)} = u_k^{(3)}$ である。n 番目のデータに対する損失関数は次式のとおりである。

$$E_n = \frac{1}{2}\left\{(z_1^{(3)} - d_1)^2 + (z_2^{(3)} - d_2)^2\right\}$$

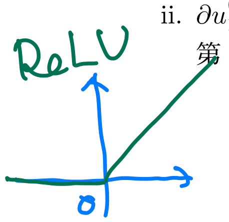
ただし、$d_1 = 4, d_2 = 2$ である。

i. $\delta_1^{(3)}$ と $\delta_2^{(3)}$ の値を求めよ。ただし、$u_1^{(3)} = 3, u_2^{(3)} = 3$ である。

$$\delta_1^{(3)} = z_1^{(3)} - d_1 = 3 - 4 = -1$$

$$\delta_2^{(3)} = z_2^{(3)} - d_2 = 3 - 2 = 1$$

ii. $\partial u_1^{(3)}/\partial u_2^{(2)}$ と $\partial u_2^{(3)}/\partial u_2^{(2)}$ の値を求めよ。ただし、$w_{12}^{(3)} = 1/2, w_{22}^{(3)} = -1/4$ であり、$u_2^{(2)} = 3/7$ である。また、第 2 層の活性化関数は ReLU 関数であり、次式のとおりである。



$$z_j^{(2)} = \begin{cases} u_j^{(2)}, & (u_j^{(2)} > 0) \\ 0, & (\text{otherwise}). \end{cases} \qquad \text{つまり} \quad f'(u_j^{(2)}) = \begin{cases} 1, & (u_j^{(2)} > 0), \\ 0, & (\text{otherwise}). \end{cases}$$

3層目から
2層目へ

$$\begin{cases} \dfrac{\partial u_1^{(3)}}{\partial u_2^{(2)}} = w_{12}^{(3)} f'(u_2^{(2)}) = \dfrac{1}{2} \cdot 1 = \dfrac{1}{2} \\ \dfrac{\partial u_2^{(3)}}{\partial u_2^{(2)}} = w_{22}^{(3)} f'(u_2^{(2)}) = -\dfrac{1}{4} \cdot 1 = -\dfrac{1}{4} \end{cases}$$

iii. $\delta_2^{(2)}$ の値を求めよ。

$$\delta_2^{(2)} = \delta_1^{(3)} \frac{\partial u_1^{(3)}}{\partial u_2^{(2)}} + \delta_2^{(3)} \frac{\partial u_2^{(3)}}{\partial u_2^{(2)}}$$

$$= -1 \cdot \frac{1}{2} + 1 \cdot \left(-\frac{1}{4}\right) = -\frac{3}{4}$$

iv. $\underbrace{\partial E_n/\partial w_{23}^{(2)}}$ の値を求めよ。ただし、$z_3^{(1)} = 4$ である。

$$= \delta_2^{(2)} \frac{\partial u_2^{(2)}}{\partial w_{23}^{(2)}} = -\frac{3}{4} \cdot z_2^{(1)}$$

$$= -\frac{3}{4} \cdot 4$$

$$= -3 \;//$$