深層学習(MLP)4章_前半

# 誤差逆伝搬法

# 4.1勾配計算の難しさ

- 勾配降下法を実行する時に必要となるパラメータ更新のための誤差関数の勾配を求める

誤差関数の勾配： $\nabla E = \dfrac{\partial E(w)}{\partial w}$

例えば二乗誤差の第l層の重みは以下の微分値で表せる

$$\frac{\partial E_n}{\partial w_{ji}^{(l)}} = (y(x_n) - d_n)^T \frac{\partial y}{\partial w_{ji}^{(l)}}$$

これだとyの中にも対象のパラメータ wが含まれているので、 $\dfrac{\partial y}{\partial w_{ji}^{(l)}}$ の部分を求める必要があるが,入力層に近い場合は以下のように出力層から順に求めると大変になる

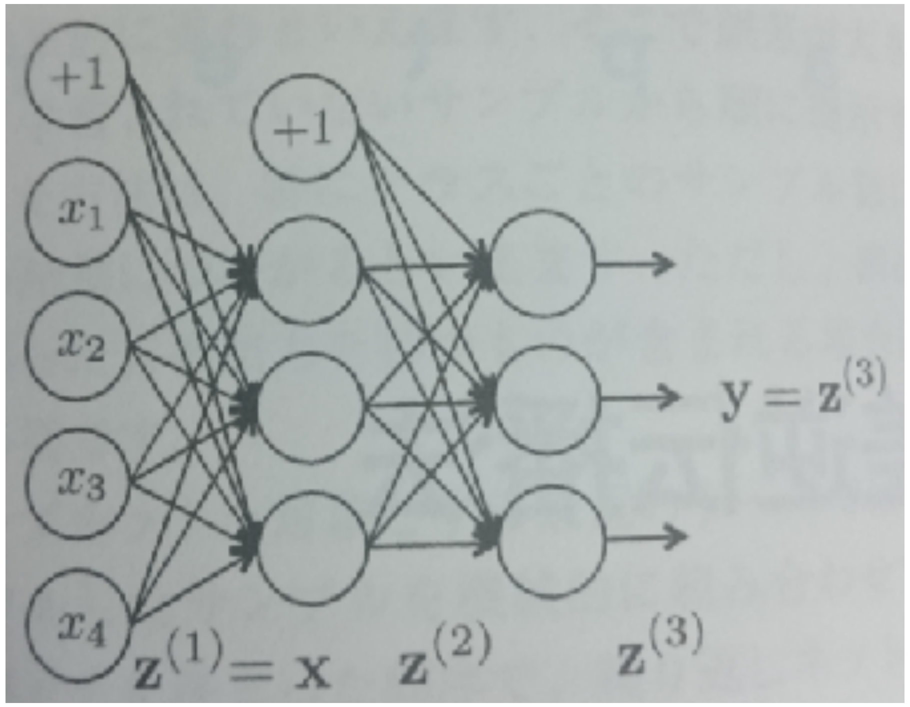
$$\begin{aligned} y(x) = f(u^{(L)}) &= f(W^L z^{(L-1)} + b^{(L)}) \\ &= f(W^{(L)} f(W^{(L-1)} z^{(L-2)} + b^{(L-1)} + b^L) \\ &= f(W^{(L)} f(W^{(L-1)} f(f(W^{(l)} z^{(l-1)} + b^{(l)}))) + b^{(L)}) \end{aligned}$$

# 4.1勾配計算の難しさ

- 層が深い場合、入力層に近い位置の誤差関数微分値を求めるのが大変
  →誤差逆伝搬ではこの問題を解決する

以降の表記についてバイアスbの表記を簡潔に行うため重みwojをbjとし、z0の出力が常に1になるようにする。

$$u_j^{(l)} = \sum_{i=1}^{n} w_{ji}^{(l)} z_i^{(l-1)} + b_j = \sum_{i=0}^{n} w_{ji}^{(l)} z_i^{(l-1)}$$



# 4.2 2層ネットワークでの計算

2層ネットワークの場合、中間層の出力は以下のようになる

$$z_j^{(2)} = f(u_j^{(2)} = f(\sum_i w_{ji}^{(2)} z_i^{(1)})$$

出力層の活性化関数は恒等写像で出力は以下になる。

$$y_j(x) = z_j^{(3)} = u_j^{(3)} = \sum_i w_{ji}^{(3)} z_i^{(2)}$$

このネットワークの誤差関数に二乗誤差を選んだ時に中間層、出力層の重みによる微分を求めてみる

# 4.2 2層ネットワークでの計算

- 出力層の重みによる微分

$$\frac{\partial E_n}{\partial w_{ji}^{(3)}} = (y(x) - d)^T \frac{\partial y}{\partial w_{ji}^{(3)}}$$

この時 $\frac{\partial y}{\partial w_{ji}^{(3)}}$ は一つの成分のみがziでそれ以外は0、つまり以下の状態である

$$\frac{\partial y}{\partial w_{ji}^{(3)}} = \left(0, , , 0, z_i^{(2)}, 0, , , , 0\right)^T$$

これより出力層の重みwjiの勾配は以下のように求められる

$$\frac{\partial E_n}{\partial w_{ji}^{(3)}} = (y_j(x) - d_j)z_i^{(2)}$$

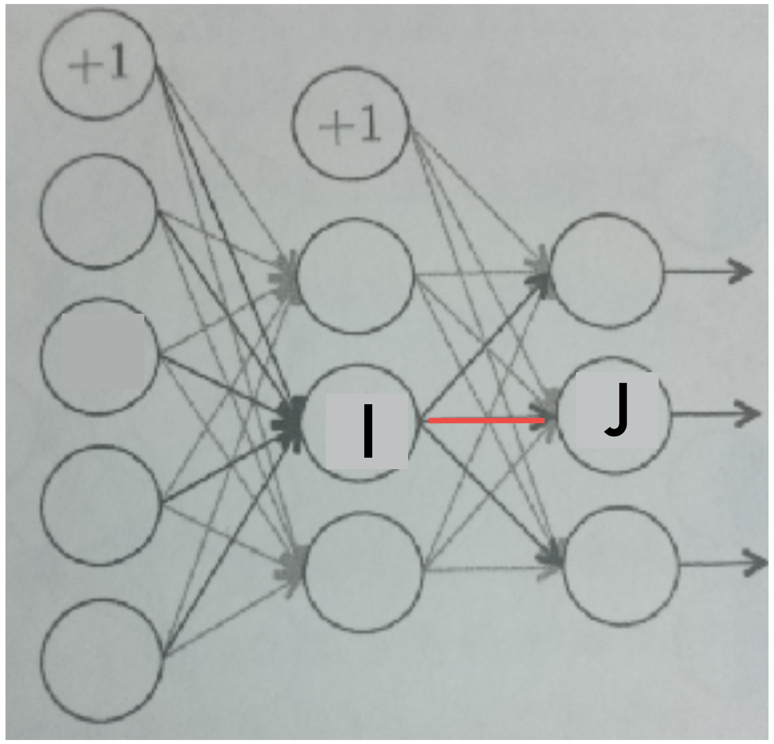
# 4.2 2層ネットワークでの計算

- 出力層の重みによる微分

$$\frac{\partial E_n}{\partial w_{ji}^{(3)}} = \boxed{(y_j(x) - d_j)}\boxed{z_i^{(2)}}$$

誤差関数になる

対象のパラメータによる出力層への入力になる



# 4.2 2層ネットワークでの計算

- 中間層の重みによる微分

ujによる偏微分で以下のように表せる

$$\frac{\partial E_n}{\partial w_{ji}^{(2)}} = \frac{\partial E_n}{\partial u_j^{(2)}} \frac{\partial u_j^{(2)}}{\partial w_{ji}^{(2)}}$$

$u_j^{(2)} = \sum_i w_{ji}^{(2)} z_i^{(1)}$ を使うと上記の右辺の第二項は以下のようになる。

$$\frac{\partial E_n}{\partial w_{ji}^{(2)}} = z_i^{(1)}$$

右辺の第一項については出力そうの各ユニットへの入力ukを使い以下のように表せる

$$\frac{\partial E_n}{\partial u_j^{(2)}} = \sum_k \frac{\partial E_n}{\partial u_k^{(3)}} \frac{\partial u_k^{(3)}}{\partial u_j^{(2)}}$$

# 4.2 2層ネットワークでの計算

- 中間層の重みによる微分

$$\frac{\partial E_n}{\partial u_j^{(2)}} = \sum_k \frac{\partial E_n}{\partial u_k^{(3)}} \frac{\partial u_k^{(3)}}{\partial u_j^{(2)}}$$

上記右辺の第一項は二乗誤差 $E_n = \frac{1}{2}\sum_k (yk(x) - d_k)^2 = \frac{1}{2}\sum_k (u_k^{(3)} - d_k)^2$
より以下のように求められる

$$\frac{\partial E_n}{\partial u_k^{(3)}} = u_k^{(3)} - d_k$$

次に第二項は $u_k^{(3)} = \sum_j w_{kj}^{(3)} f(u_j^{(2)})$ より以下のように求められる

$$\frac{\partial u_k^{(3)}}{\partial u_j^{(2)}} = w_{kj}^{(3)} f'(u_j^{(2)})$$

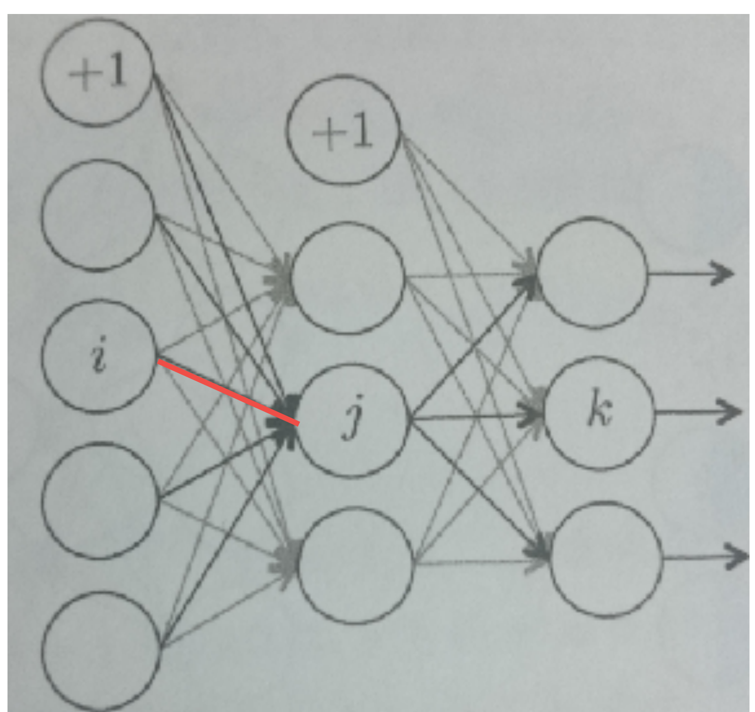
# 4.2 2層ネットワークでの計算

- 中間層の重みによる微分

最終的に中間層の重みwjiは以下のように求められる

$$\frac{\partial E_n}{\partial w_{ji}^{(2)}} = \left( \boxed{f'(u_j^{(2)})} \boxed{\sum_k w_{kj}^{(3)} (u_k^{(3)} - d_k)} \right)$$

**入力層側から計算する**　**出力層側から計算する**



# 4.2 2層ネットワークでの計算

- 中間層の重みによる微分

最終的に中間層の重みwjiは以下のように求められる

$$\frac{\partial E_n}{\partial w_{ji}^{(2)}} = \left( \boxed{f'(u_j^{(2)})} \boxed{\sum_k w_{kj}^{(3)} (u_k^{(3)} - d_k)} \right)$$

**入力層側から計算する**　**出力層側から計算する**

入力層側からの計算と出力層側で独立した計算により微分ができるようになっている