機械学習（5回目）

創域理工学部　情報計算科学科
桂田　浩一

1

10/5/2023

前回の復習（１）

- 誤差逆伝播法
  - 上位層から下位層に向かって（ニューラルネットの入出力と方向に）誤差関数の重み，閾値に対する偏微分と更新値を求めていく方法
- これからの説明でのネットワーク設定
  - 誤差関数：　二乗誤差（$E = \frac{1}{2}\sum_{j=1}^{J}\left(t_j - o_j^{(L)}\right)^2$）
  - 活性化関数：　$f(x)$
  - 層の数：　L
  - 各ノードの出力：　$o_i^{(l)}$（第 l 層の i 番目のノードの出力）
  - 結合荷重：　$w_{j,i}^{(l-1)}$
（第 $l-1$ 層の i 番目のノードと第 l 層の j 番目のノードの間のリンク）

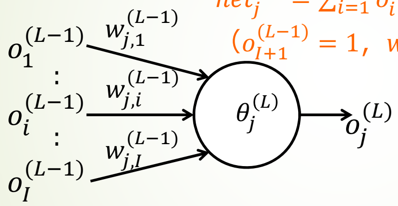
前回の復習（2）

- 出力層（第 L 層）とその一つ前の層（第 $L-1$ 層）の間の重み $w_{j,i}^{(L-1)}$ について

$net_j^{(L)} = \sum_{i=1}^{I+1} o_i^{(L-1)} w_{j,i}^{(L-1)}$

$(o_{I+1}^{(L-1)} = 1,\ w_{j,I+1}^{(L-1)} = -\theta_j^{(L)})$

$o_j^{(L)} = f\left(net_j^{(L)}\right)$



t_j：教師データ

$$E = \frac{1}{2}\sum_{j=1}^{J}\left(t_j - o_j^{(L)}\right)^2$$

$$\frac{\partial E}{\partial w_{j,i}^{(L-1)}} = \frac{\partial E}{\partial net_j^{(L)}} \underbrace{\boxed{\frac{\partial net_j^{(L)}}{\partial w_{j,i}^{(L-1)}}}}_{= o_i^{(L-1)}}$$

←チェーンルール

（ノード j の $net_j^{(L)}$ 以外が $w_{j,i}^{(L-1)}$ に依存しないことに注意！！）

前回の復習（3）

- 出力層（第 L 層）とその一つ前の層（第 $L-1$ 層）の間の重み $w_{j,i}^{(L-1)}$ について

$net_j^{(L)} = \sum_{i=1}^{I+1} o_i^{(L-1)} w_{j,i}^{(L-1)}$

$(o_{I+1}^{(L-1)} = 1,\ w_{j,I+1}^{(L-1)} = -\theta_j^{(L)})$

$o_j^{(L)} = f\left(net_j^{(L)}\right)$

$E = \frac{1}{2}\sum_{j=1}^{J}\left(t_j - o_j^{(L)}\right)^2$

$$\frac{\partial E}{\partial w_{j,i}^{(L-1)}} = \boxed{\frac{\partial E}{\partial net_j^{(L)}}}\ \boxed{\frac{\partial net_j^{(L)}}{\partial w_{j,i}^{(L-1)}}} = o_i^{(L-1)}$$

$$\boxed{\frac{\partial E}{\partial net_j^{(L)}}} = \underbrace{\boxed{\frac{\partial E}{\partial o_j^{(L)}}}}_{= -\left(t_j - o_j^{(L)}\right)}\ \boxed{\frac{\partial o_j^{(L)}}{\partial net_j^{(L)}}} = f'\left(net_j^{(L)}\right)$$

前回の復習（4）

- 出力層（第 L 層）とその一つ前の層（第 $L-1$ 層）の間の重み $w_{j,i}^{(L-1)}$ について

$$\frac{\partial E}{\partial w_{j,i}^{(L-1)}} = \boxed{\frac{\partial E}{\partial net_j^{(L)}}} \boxed{\frac{\partial net_j^{(L)}}{\partial w_{j,i}^{(L-1)}}} \quad o_i^{(L-1)}$$

$$\boxed{\frac{\partial E}{\partial net_j^{(L)}}} = \boxed{\frac{\partial E}{\partial o_j^{(L)}}} \boxed{\frac{\partial o_j^{(L)}}{\partial net_j^{(L)}}} \quad f'\left(net_j^{(L)}\right) \quad -\left(t_j - o_j^{(L)}\right)$$

$$= -f'\left(net_j^{(L)}\right)\left(t_j - o_j^{(L)}\right)$$

ここで $f'\left(net_j^{(L)}\right)\left(t_j - o_j^{(L)}\right) = \delta_j^{(L)}$ と置くと $\boxed{\frac{\partial E}{\partial net_j^{(L)}}} = -\delta_j^{(L)}$

$$\frac{\partial E}{\partial w_{j,i}^{(L-1)}} = -\delta_j^{(L)} o_i^{(L-1)}$$

本日の内容

- 誤差逆伝播法の続き
  - 中間層の偏微分の求め方
  - 誤差逆伝播法アルゴリズム

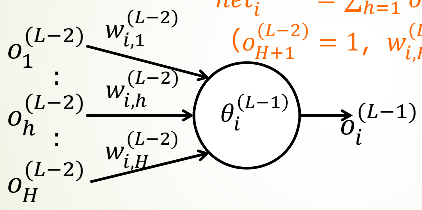
誤差逆伝播法（5）

- 第 $L-1$ 層とその一つ前の層（第 $L-2$ 層）の間の重み $w_{i,h}^{(L-2)}$ について

$$net_i^{(L-1)} = \sum_{h=1}^{H+1} o_h^{(L-2)} w_{i,h}^{(L-2)}$$

$$(o_{H+1}^{(L-2)} = 1,\ w_{i,H+1}^{(L-2)} = -\theta_i^{(L-1)})$$

$$o_i^{(L-1)} = f\left(net_i^{(L-1)}\right)$$



$$E = \frac{1}{2}\sum_{j=1}^{J}\left(t_j - o_j^{(L)}\right)^2$$

$$\frac{\partial E}{\partial w_{i,h}^{(L-2)}} = \frac{\partial E}{\partial net_i^{(L-1)}} \underbrace{\boxed{\frac{\partial net_i^{(L-1)}}{\partial w_{i,h}^{(L-2)}}}}_{= o_h^{(L-2)}}$$ ←チェーンルール

（ノード i の $net_i^{(L-1)}$ 以外が $w_{i,h}^{(L-2)}$ に依存しないことに注意！！）

誤差逆伝播法（6）

- 第 $L-1$ 層とその一つ前の層（第 $L-2$ 層）の間の重み $w_{i,h}^{(L-2)}$ について

$$net_i^{(L-1)} = \sum_{h=1}^{H+1} o_h^{(L-2)} w_{i,h}^{(L-2)}$$

$$(o_{H+1}^{(L-2)} = 1,\ w_{i,H+1}^{(L-2)} = -\theta_i^{(L-1)})$$

$$o_i^{(L-1)} = f\left(net_i^{(L-1)}\right)$$

$$E = \frac{1}{2}\sum_{j=1}^{J}\left(t_j - o_j^{(L)}\right)^2$$

$$\frac{\partial E}{\partial w_{i,h}^{(L-2)}} = \boxed{\frac{\partial E}{\partial net_i^{(L-1)}}}\ \boxed{\frac{\partial net_i^{(L-1)}}{\partial w_{i,h}^{(L-2)}}} = o_h^{(L-2)}$$

$$\boxed{\frac{\partial E}{\partial net_i^{(L-1)}}} = \underbrace{\boxed{\frac{\partial E}{\partial o_i^{(L-1)}}}}_{= ?}\ \boxed{\frac{\partial o_i^{(L-1)}}{\partial net_i^{(L-1)}}} = f'\left(net_i^{(L-1)}\right)$$

誤差逆伝播法（7）

- 第 $L-1$ 層とその一つ前の層（第 $L-2$ 層）の間の重み $w_{i,h}^{(L-2)}$ について

$$net_i^{(L-1)} = \sum_{h=1}^{H+1} o_h^{(L-2)} w_{i,h}^{(L-2)}$$
$$(o_{H+1}^{(L-2)} = 1,\ w_{i,H+1}^{(L-2)} = -\theta_i^{(L-1)})$$
$$o_i^{(L-1)} = f\left(net_i^{(L-1)}\right)$$
$$E = \frac{1}{2}\sum_{j=1}^{J}\left(t_j - o_j^{(L)}\right)^2$$
$$net_j^{(L)} = \sum_{i=1}^{I+1} o_i^{(L-1)} w_{j,i}^{(L-1)}$$
$$(o_{I+1}^{(L-1)} = 1,\ w_{j,I+1}^{(L-1)} = -\theta_j^{(L)})$$

$$\frac{\partial E}{\partial w_{i,h}^{(L-2)}} = \boxed{\frac{\partial E}{\partial net_i^{(L-1)}}}\boxed{\frac{\partial net_i^{(L-1)}}{\partial w_{i,h}^{(L-2)}}} = o_h^{(L-2)}$$

$$\boxed{\frac{\partial E}{\partial net_i^{(L-1)}}} = \boxed{\frac{\partial E}{\partial o_i^{(L-1)}}}\boxed{\frac{\partial o_i^{(L-1)}}{\partial net_i^{(L-1)}}} = f'\left(net_i^{(L-1)}\right)$$

$$\boxed{\frac{\partial E}{\partial o_i^{(L-1)}}} = \sum_j \underbrace{\boxed{\frac{\partial E}{\partial net_j^{(L)}}}}_{= -\delta_j^{(L)}}\underbrace{\boxed{\frac{\partial net_j^{(L)}}{\partial o_i^{(L-1)}}}}_{= w_{j,i}^{(L-1)}}$$

←チェーンルール

（全ての $net_j^{(L)}$ が $o_i^{(L-1)}$ に依存する！！）

誤差逆伝播法（8）

- 第 $L-1$ 層とその一つ前の層（第 $L-2$ 層）の間の重み $w_{i,h}^{(L-2)}$ について

$$net_i^{(L-1)} = \sum_{h=1}^{H+1} o_h^{(L-2)} w_{i,h}^{(L-2)}$$
$$(o_{H+1}^{(L-2)} = 1,\ w_{i,H+1}^{(L-2)} = -\theta_i^{(L-1)})$$
$$o_i^{(L-1)} = f\left(net_i^{(L-1)}\right)$$
$$E = \frac{1}{2}\sum_{j=1}^{J}\left(t_j - o_j^{(L)}\right)^2$$
$$net_j^{(L)} = \sum_{i=1}^{I+1} o_i^{(L-1)} w_{j,i}^{(L-1)}$$
$$(o_{I+1}^{(L-1)} = 1,\ w_{j,I+1}^{(L-1)} = -\theta_j^{(L)})$$

$$\frac{\partial E}{\partial w_{i,h}^{(L-2)}} = \boxed{\frac{\partial E}{\partial net_i^{(L-1)}}}\boxed{\frac{\partial net_i^{(L-1)}}{\partial w_{i,h}^{(L-2)}}} = o_h^{(L-2)}$$

$$\boxed{\frac{\partial E}{\partial net_i^{(L-1)}}} = \boxed{\frac{\partial E}{\partial o_i^{(L-1)}}}\boxed{\frac{\partial o_i^{(L-1)}}{\partial net_i^{(L-1)}}} = f'\left(net_i^{(L-1)}\right)$$

$$\boxed{\frac{\partial E}{\partial o_i^{(L-1)}}} = \sum_j \underbrace{\boxed{\frac{\partial E}{\partial net_j^{(L)}}}}_{= -\delta_j^{(L)}}\underbrace{\boxed{\frac{\partial net_j^{(L)}}{\partial o_i^{(L-1)}}}}_{= w_{j,i}^{(L-1)}} = -\sum_j \delta_j^{(L)} w_{j,i}^{(L-1)}$$

誤差逆伝播法（９）

- 第 $L-1$ 層とその一つ前の層（第 $L-2$ 層）の間の重み $w_{i,h}^{(L-2)}$ について

$$\frac{\partial E}{\partial w_{i,h}^{(L-2)}} = \boxed{\frac{\partial E}{\partial net_i^{(L-1)}}}\boxed{\frac{\partial net_i^{(L-1)}}{\partial w_{i,h}^{(L-2)}}} = -\delta_i^{(L-1)} o_h^{(L-2)}$$

$\delta_i^{(L-1)}$ と置くと

$$\boxed{\frac{\partial E}{\partial net_i^{(L-1)}}} = \boxed{\frac{\partial E}{\partial o_i^{(L-1)}}}\boxed{\frac{\partial o_i^{(L-1)}}{\partial net_i^{(L-1)}}} = -\overset{\delta_i^{(L-1)}}{\overset{\parallel}{\boxed{f'\left(net_i^{(L-1)}\right)\sum_j \delta_j^{(L)} w_{j,i}^{(L-1)}}}}$$

$$\boxed{\frac{\partial E}{\partial o_i^{(L-1)}}} = \sum_j \underset{-\delta_j^{(L)}}{\underset{\parallel}{\boxed{\frac{\partial E}{\partial net_j^{(L)}}}}}\underset{w_{j,i}^{(L-1)}}{\underset{\parallel}{\boxed{\frac{\partial net_j^{(L)}}{\partial o_i^{(L-1)}}}}} = -\sum_j \delta_j^{(L)} w_{j,i}^{(L-1)}$$

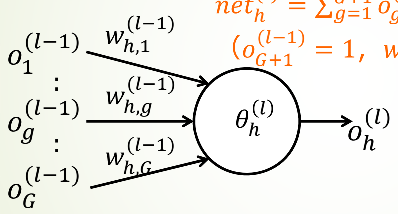
誤差逆伝播法（１０）

- 中間層の第 l 層とその一つ前の層（第 $l-1$ 層）の間の重み $w_{h,g}^{(l-1)}$ について

$net_h^{(l)} = \sum_{g=1}^{G+1} o_g^{(l-1)} w_{h,g}^{(l-1)}$

$(o_{G+1}^{(l-1)} = 1,\ w_{h,G+1}^{(l-1)} = -\theta_h^{(l)})$

$o_h^{(l)} = f\left(net_h^{(l)}\right)$



$$E = \frac{1}{2}\sum_{j=1}^{J}\left(t_j - o_j^{(L)}\right)^2$$

$$\frac{\partial E}{\partial w_{h,g}^{(l-1)}} = \frac{\partial E}{\partial net_h^{(l)}}\underset{o_g^{(l-1)}}{\underset{\parallel}{\boxed{\frac{\partial net_h^{(l)}}{\partial w_{h,g}^{(l-1)}}}}}$$

←チェーンルール

（ノード h の $net_h^{(l)}$ 以外が $w_{h,g}^{(l-1)}$ に依存しないことに注意！！）

誤差逆伝播法（１１）

- 中間層の第 l 層とその一つ前の層（第 $l-1$ 層）の間の重み $w_{h,g}^{(l-1)}$ について

$net_h^{(l)} = \sum_{g=1}^{G+1} o_g^{(l-1)} w_{h,g}^{(l-1)}$
$(o_{G+1}^{(l-1)} = 1,\ w_{h,G+1}^{(l-1)} = -\theta_h^{(l)})$

$o_h^{(l)} = f\left(net_h^{(l)}\right)$

$E = \frac{1}{2}\sum_{j=1}^{J}\left(t_j - o_j^{(L)}\right)^2$

$$\frac{\partial E}{\partial w_{h,g}^{(l-1)}} = \boxed{\frac{\partial E}{\partial net_h^{(l)}}}\boxed{\frac{\partial net_h^{(l)}}{\partial w_{h,g}^{(l-1)}}} = o_g^{(l-1)}$$

$$\boxed{\frac{\partial E}{\partial net_h^{(l)}}} = \underbrace{\boxed{\frac{\partial E}{\partial o_h^{(l)}}}}_{=\,?}\boxed{\frac{\partial o_h^{(l)}}{\partial net_h^{(l)}}} = f'\left(net_h^{(l)}\right)$$

誤差逆伝播法（１２）

- 中間層の第 l 層とその一つ前の層（第 $l-1$ 層）の間の重み $w_{h,g}^{(l-1)}$ について

$net_h^{(l)} = \sum_{g=1}^{G+1} o_g^{(l-1)} w_{h,g}^{(l-1)}$
$(o_{G+1}^{(l-1)} = 1,\ w_{h,G+1}^{(l-1)} = -\theta_h^{(l)})$

$E = \frac{1}{2}\sum_{j=1}^{J}\left(t_j - o_j^{(L)}\right)^2$

$net_i^{(l+1)} = \sum_{h=1}^{H+1} o_h^{(l)} w_{i,h}^{(l)}$
$(o_{H+1}^{(l)} = 1,\ w_{i,H+1}^{(l)} = -\theta_i^{(l+1)})$

$$\frac{\partial E}{\partial w_{h,g}^{(l-1)}} = \boxed{\frac{\partial E}{\partial net_h^{(l)}}}\boxed{\frac{\partial net_h^{(l)}}{\partial w_{h,g}^{(l-1)}}} = o_g^{(l-1)}$$

$$\boxed{\frac{\partial E}{\partial net_h^{(l)}}} = \boxed{\frac{\partial E}{\partial o_h^{(l)}}}\boxed{\frac{\partial o_h^{(l)}}{\partial net_h^{(l)}}} = f'\left(net_h^{(l)}\right)$$

$$\boxed{\frac{\partial E}{\partial o_h^{(l)}}} = \sum_i \underbrace{\boxed{\frac{\partial E}{\partial net_i^{(l+1)}}}}_{=\,-\delta_i^{(l+1)}}\underbrace{\boxed{\frac{\partial net_i^{(l+1)}}{\partial o_h^{(l)}}}}_{=\,w_{i,h}^{(l)}}$$

←チェーンルール
（全ての $net_i^{(l+1)}$ が $o_h^{(l)}$ に依存する！！）

誤差逆伝播法（１３）

- 中間層の第 l 層とその一つ前の層（第 $l-1$ 層）の間の重み $w_{h,g}^{(l-1)}$ について

$$net_h^{(l)} = \sum_{g=1}^{G+1} o_g^{(l-1)} w_{h,g}^{(l-1)}$$
$$(o_{G+1}^{(l-1)} = 1,\ w_{h,G+1}^{(l-1)} = -\theta_h^{(l)})$$

$$\frac{\partial E}{\partial w_{h,g}^{(l-1)}} = \boxed{\frac{\partial E}{\partial net_h^{(l)}}}\ \boxed{\frac{\partial net_h^{(l)}}{\partial w_{h,g}^{(l-1)}}} = o_g^{(l-1)}$$

$$E = \frac{1}{2}\sum_{j=1}^{J}\left(t_j - o_j^{(L)}\right)^2$$

$$\boxed{\frac{\partial E}{\partial net_h^{(l)}}} = \boxed{\frac{\partial E}{\partial o_h^{(l)}}}\ \boxed{\frac{\partial o_h^{(l)}}{\partial net_h^{(l)}}} = f'\left(net_h^{(l)}\right)$$

$$net_i^{(l+1)} = \sum_{h=1}^{H+1} o_h^{(l)} w_{i,h}^{(l)}$$
$$(o_{H+1}^{(l)} = 1,\ w_{i,H+1}^{(l)} = -\theta_i^{(l+1)})$$

$$\boxed{\frac{\partial E}{\partial o_h^{(l)}}} = \sum_i \underbrace{\boxed{\frac{\partial E}{\partial net_i^{(l+1)}}}}_{-\delta_i^{(l+1)}}\ \underbrace{\boxed{\frac{\partial net_i^{(l+1)}}{\partial o_h^{(l)}}}}_{w_{i,h}^{(l)}} = -\sum_i \delta_i^{(l+1)} w_{i,h}^{(l)}$$

誤差逆伝播法（１４）

- 中間層の第 l 層とその一つ前の層（第 $l-1$ 層）の間の重み $w_{h,g}^{(l-1)}$ について

$$\frac{\partial E}{\partial w_{h,g}^{(l-1)}} = \boxed{\frac{\partial E}{\partial net_h^{(l)}}}\ \boxed{\frac{\partial net_h^{(l)}}{\partial w_{h,g}^{(l-1)}}} = -\delta_h^{(l)} o_g^{(l-1)}$$

$\delta_h^{(l)}$ と置くと

$$\boxed{\frac{\partial E}{\partial net_h^{(l)}}} = \boxed{\frac{\partial E}{\partial o_h^{(l)}}}\ \boxed{\frac{\partial o_h^{(l)}}{\partial net_h^{(l)}}} = -\overbrace{\boxed{f'\left(net_h^{(l)}\right)\sum_i \delta_i^{(l+1)} w_{i,h}^{(l)}}}^{\delta_h^{(l)}}$$

$$\boxed{\frac{\partial E}{\partial o_h^{(l)}}} = \sum_i \underbrace{\boxed{\frac{\partial E}{\partial net_i^{(l+1)}}}}_{-\delta_i^{(l+1)}}\ \underbrace{\boxed{\frac{\partial net_i^{(l+1)}}{\partial o_h^{(l)}}}}_{w_{i,h}^{(l)}} = -\sum_i \delta_i^{(l+1)} w_{i,h}^{(l)}$$

誤差逆伝播法（アルゴリズム（１））

1. 結合荷重 $w_{j,i}^{(l)}$ $(i = 1, \cdots, n(l) + 1,$
$j = 1, \cdots, n(l+1),\quad l = 1, \cdots L-1)$ の初期値をランダムに小さい値に設定する．
ここで $w_{j,i}^{(l)}$ は第 l 層と第 $l+1$ 層の間の結合荷重，
$n(l)$ は第 l 層のユニット数，L は層の数を表す．
また，$w_{j,n(l)+1}^{(l)} = -\theta_j^{(l+1)}$（第 $l+1$ 層のユニット j の閾値の符号を逆転したもの）とする．
さらに学習率 η $(0 < \eta \leq 1)$ を小さな値に設定する．

誤差逆伝播法（アルゴリズム（２））

2. 入力パターンベクトル $\boldsymbol{x}_p = (x_{p,1}, \cdots, x_{p,n(1)})$ と，それに対応する教師信号 $\boldsymbol{t}_p = (t_{p,1}, \cdots, t_{p,n(L)})$ の組を選ぶ．

3. 結合荷重 $w_{j,i}^{(l)}$ と入力パターンベクトル $\boldsymbol{x}_p$ を用いて，入力層から出力層に向けて，各ユニットの出力値を次式で計算する．ただし $net_{p,j}^{(l)} = \sum_{i=1}^{n(l-1)+1} w_{j,i}^{(l-1)} o_{p,i}^{(l-1)}$ とする．

$$o_{p,j}^{(1)} = x_{p,j} \qquad (j = 1 \cdots n(1))$$
$$o_{p,j}^{(l)} = f\left(net_{p,j}^{(l)}\right) \qquad (j = 1 \cdots n(l), 2 \leq l \leq L)$$
$$o_{p,n(l)+1}^{(l)} = 1 \qquad (1 \leq l \leq L-1)$$

f：活性化関数

誤差逆伝播法（アルゴリズム（３））

4. 出力層から入力層に向けて誤差 $\delta_{p,j}^{(l)}$ を次式に従って計算する．

E：誤差関数

$$\delta_{p,j}^{(L)} = -f'\left(net_{p,j}^{(L)}\right)\frac{\partial E}{\partial o_{p,j}^{(L)}} \quad (j = 1 \cdots n(L))$$

$$\delta_{p,j}^{(l)} = f'\left(net_{p,j}^{(l)}\right)\sum_{k=1}^{n(l+1)} \delta_{p,k}^{(l+1)} w_{k,j}^{(l)}$$

$$(j = 1 \cdots n(l),\ \ l = 2 \cdots L-1)$$

層ごとに f を変えたり E を変更したりしても他層に影響しない

5. 結合荷重を修正する．

$$w_{j,i}^{(l)} \leftarrow w_{j,i}^{(l)} + \eta\delta_{p,j}^{(l+1)} o_{p,i}^{(l)}$$

$$(i = 1 \cdots n(l)+1,\ j = 1 \cdots n(l+1),\ l = 1 \cdots L-1)$$

誤差逆伝播法（アルゴリズム（４））

6. 学習パターンに対して誤差 E が設定値以下になれば終了．そうでなければ2.~5.を繰り返す．

出題予定の演習課題

- 誤差逆伝播法について