機械学習（6回目）

創域理工学部　情報計算科学科
桂田　浩一

1

10/19/2023

前回の復習

- 誤差逆伝播法

3

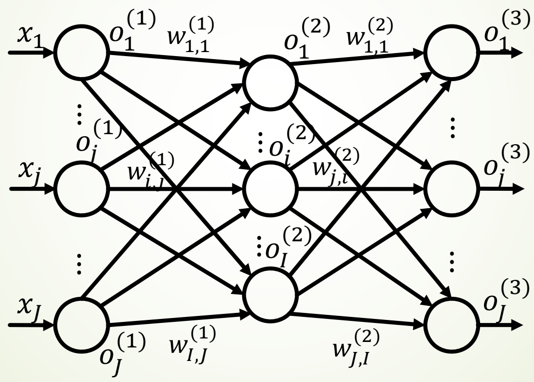
本日の内容

- 複雑な依存関係がある場合（誤差逆伝播法がそのまま適用できない場合）の偏微分の求め方

4

AutoEncoder (AE) 自己符号化器

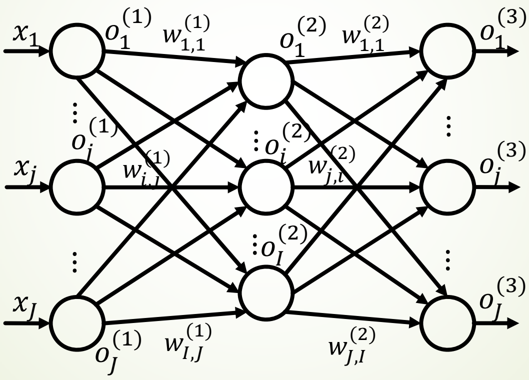
- 入力＝出力となる恒等変換を学習するニューラルネット



5

Tied Weight AutoEncoder

- $w_{i,j}^{(1)} = w_{j,i}^{(2)}$という制約を加えたAutoEncoder



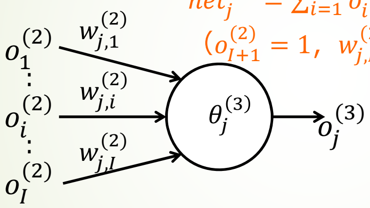
Tied Weight AEの偏微分（１）

- 出力層（第３層）とその一つ前の層（第２層）の間の重み $w_{j,i}^{(2)} = w_{i,j}^{(1)} = w_{j,i}$ $(1 \le i \le I, 1 \le j \le J)$について

$net_j^{(3)} = \sum_{i=1}^{I+1} o_i^{(2)} w_{j,i}^{(2)}$

$(o_{I+1}^{(2)} = 1,\ w_{j,I+1}^{(2)} = -\theta_j^{(3)})$

$o_j^{(3)} = f\left(net_j^{(3)}\right)$



t_j：教師データ

$\cancel{\dfrac{\partial E}{\partial w_{j,i}} = \dfrac{\partial E}{\partial net_j^{(3)}} \dfrac{\partial net_j^{(3)}}{\partial w_{j,i}}}$ ←チェーンルール？

（ノード j の $net_j^{(3)}$ 以外も $w_{i,j}^{(1)} = w_{j,i}$ に依存する！！）

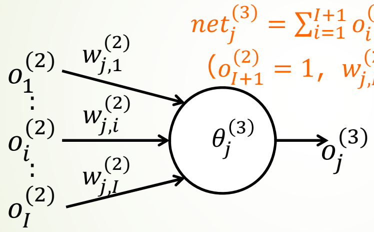
Tied Weight AEの偏微分（１）

- 出力層（第３層）とその一つ前の層（第２層）の間の重み $w_{j,i}^{(2)} = w_{i,j}^{(1)} = w_{j,i}$ $(1 \le i \le I, 1 \le j \le J)$について

$net_j^{(3)} = \sum_{i=1}^{I+1} o_i^{(2)} w_{j,i}^{(2)}$

$(o_{I+1}^{(2)} = 1,\ w_{j,I+1}^{(2)} = -\theta_j^{(3)})$

$o_j^{(3)} = f\left(net_j^{(3)}\right)$



t_j：教師データ

$$\frac{\partial E}{\partial w_{j,i}} = \sum_{k=1}^{J} \frac{\partial E}{\partial net_k^{(3)}} \frac{\partial net_k^{(3)}}{\partial w_{j,i}}$$ ←チェーンルール

（ノード j の $net_k^{(3)}$ 以外も
$w_{i,j}^{(1)} = w_{j,i}$ に依存する！！）

Tied Weight AEの偏微分（２）

- 出力層（第３層）とその一つ前の層（第２層）の間の重み $w_{j,i}^{(2)} = w_{i,j}^{(1)} = w_{j,i}$ $(1 \le i \le I, 1 \le j \le J)$について

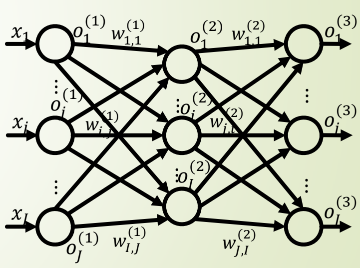
$net_j^{(3)} = \sum_{i=1}^{I+1} o_i^{(2)} w_{j,i}^{(2)}$

$(o_{I+1}^{(2)} = 1,\ w_{j,I+1}^{(2)} = -\theta_j^{(3)})$

$o_j^{(3)} = f\left(net_j^{(3)}\right)$

$$\frac{\partial E}{\partial w_{j,i}} = \sum_{k=1}^{J} \boxed{\frac{\partial E}{\partial net_k^{(3)}}} \frac{\partial net_k^{(3)}}{\partial w_{j,i}}$$

$$\boxed{\frac{\partial E}{\partial net_k^{(3)}}} = \frac{\partial E}{\partial o_k^{(3)}} \frac{\partial o_k^{(3)}}{\partial net_k^{(3)}} = \frac{\partial E}{\partial o_k^{(3)}} f'\left(net_k^{(3)}\right)$$



Tied Weight AEの偏微分（３）

- 出力層（第３層）とその一つ前の層（第２層）の間の重み $w_{j,i}^{(2)} = w_{i,j}^{(1)} = w_{j,i}$ $(1 \le i \le I, 1 \le j \le J)$について

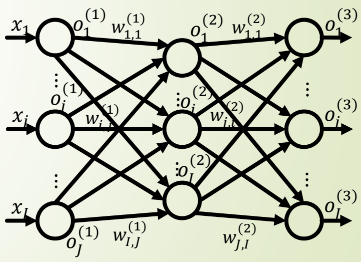
$$net_j^{(3)} = \sum_{i=1}^{I+1} o_i^{(2)} w_{j,i}^{(2)} \qquad o_j^{(3)} = f\left(net_j^{(3)}\right)$$

$$\left(o_{I+1}^{(2)} = 1,\ w_{j,I+1}^{(2)} = -\theta_j^{(3)}\right)$$

$$\frac{\partial E}{\partial w_{j,i}} = \sum_{k=1}^{J} \frac{\partial E}{\partial net_k^{(3)}} \boxed{\frac{\partial net_k^{(3)}}{\partial w_{j,i}}}$$

$$\boxed{\frac{\partial net_k^{(3)}}{\partial w_{j,i}}} = \frac{\partial\left(\sum_{h=1}^{I+1} o_h^{(2)} w_{k,h}^{(2)}\right)}{\partial w_{j,i}} = \boxed{\frac{\partial\left(o_i^{(2)} w_{k,i}^{(2)}\right)}{\partial w_{j,i}}}$$

$h \ne i$ なら $o_h^{(2)}$ も $w_{k,h}^{(2)}$ も $w_{j,i}$ に非依存



Tied Weight AEの偏微分（４）

- 出力層（第３層）とその一つ前の層（第２層）の間の重み $w_{j,i}^{(2)} = w_{i,j}^{(1)} = w_{j,i}$ $(1 \le i \le I, 1 \le j \le J)$について

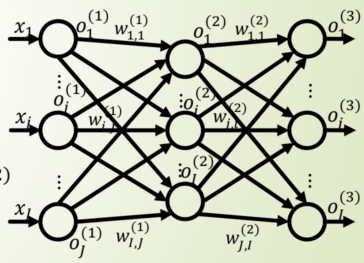
$$\boxed{\frac{\partial net_k^{(3)}}{\partial w_{j,i}}} = \frac{\partial\left(\sum_{h=1}^{I+1} o_h^{(2)} w_{k,h}^{(2)}\right)}{\partial w_{j,i}} = \boxed{\frac{\partial\left(o_i^{(2)} w_{k,i}^{(2)}\right)}{\partial w_{j,i}}}$$

$k = j$ のとき $o_i^{(2)}$ も $w_{k,i}^{(2)}$ も $w_{j,i}$ に依存

$$\frac{\partial\left(o_i^{(2)} w_{k,i}^{(2)}\right)}{\partial w_{j,i}} = \frac{\partial o_i^{(2)}}{\partial w_{j,i}} w_{k,i}^{(2)} + \frac{\partial w_{k,i}^{(2)}}{\partial w_{j,i}} o_i^{(2)} = \boxed{\frac{\partial o_i^{(2)}}{\partial w_{j,i}}} w_{k,i}^{(2)} + o_i^{(2)}$$

$k \ne j$ のとき $o_i^{(2)}$ のみが $w_{j,i}$ に依存

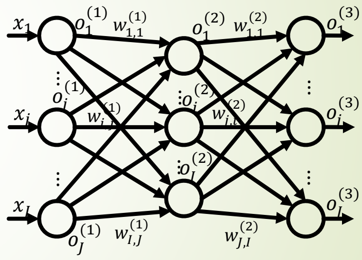
$$\frac{\partial\left(o_i^{(2)} w_{k,i}^{(2)}\right)}{\partial w_{j,i}} = \frac{\partial o_i^{(2)}}{\partial w_{j,i}} w_{k,i}^{(2)} = \boxed{\frac{\partial o_i^{(2)}}{\partial w_{j,i}}} w_{k,i}^{(2)}$$



Tied Weight AEの偏微分（５）

- 出力層（第３層）とその一つ前の層（第２層）の間の重み $w_{j,i}^{(2)} = w_{i,j}^{(1)} = w_{j,i}$ $(1 \le i \le I, 1 \le j \le J)$について

$$\boxed{\frac{\partial o_i^{(2)}}{\partial w_{j,i}}} = \underbrace{\boxed{\frac{\partial o_i^{(2)}}{\partial net_i^{(2)}}}}_{f'\left(net_i^{(2)}\right)} \underbrace{\boxed{\frac{\partial net_i^{(2)}}{\partial w_{j,i}}}}_{o_j^{(1)}}$$



$$\frac{\partial E}{\partial w_{j,i}} = \sum_{k=1}^{J} \frac{\partial E}{\partial net_k^{(3)}} \frac{\partial net_k^{(3)}}{\partial w_{j,i}}$$

$$= \sum_{k=1}^{J} \frac{\partial E}{\partial o_k^{(3)}} f'\left(net_k^{(3)}\right) w_{k,i}^{(2)} f'\left(net_i^{(2)}\right) o_j^{(1)}$$

$$+ \frac{\partial E}{\partial o_j^{(3)}} f'\left(net_j^{(3)}\right) o_i^{(2)}$$

$net_i^{(2)} = \sum_{j=1}^{J+1} o_j^{(1)} w_{i,j}^{(1)}$

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

$o_i^{(2)} = f\left(net_i^{(2)}\right)$

出題予定の演習課題

- 依存関係のあるネットワークの偏微分