計算知能 (COMPUTATIONAL INTELLIGENCE)

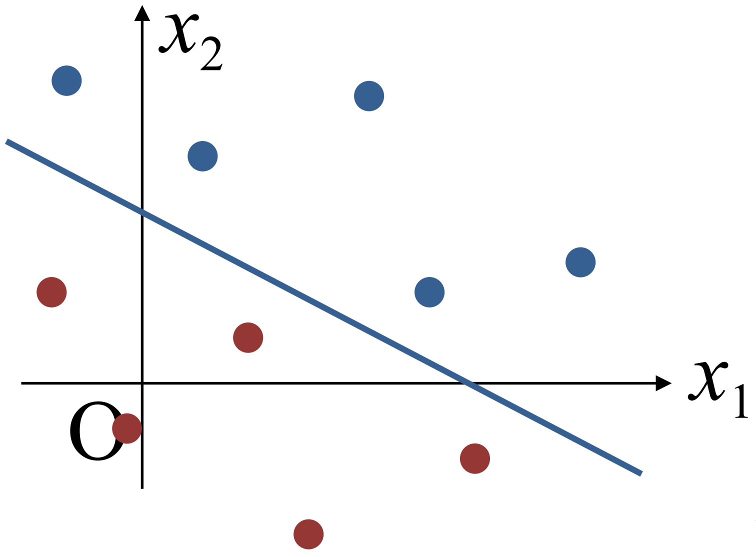
第７回　多層ニューラルネットワーク

教員：　谷口彰

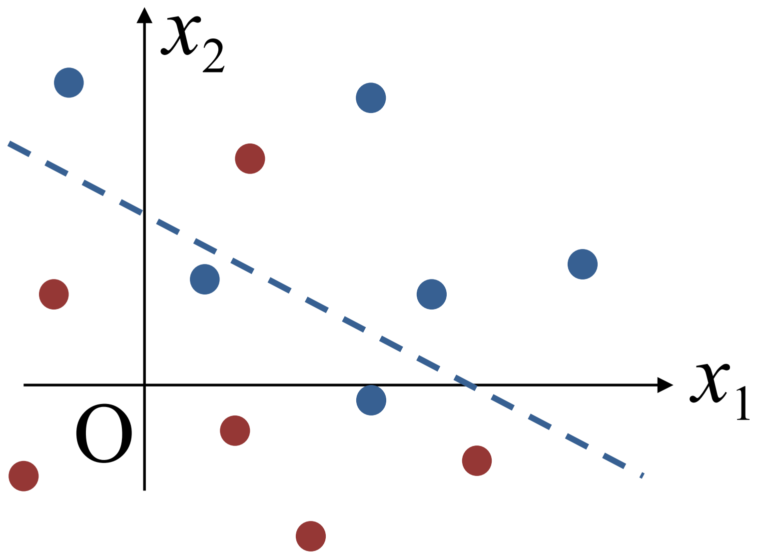
今回の内容

- シグモイドの３層ネットワーク
  - 一般化デルタ則

復習：線形分離可能な2次元データ



線形分離不可能な2次元データ

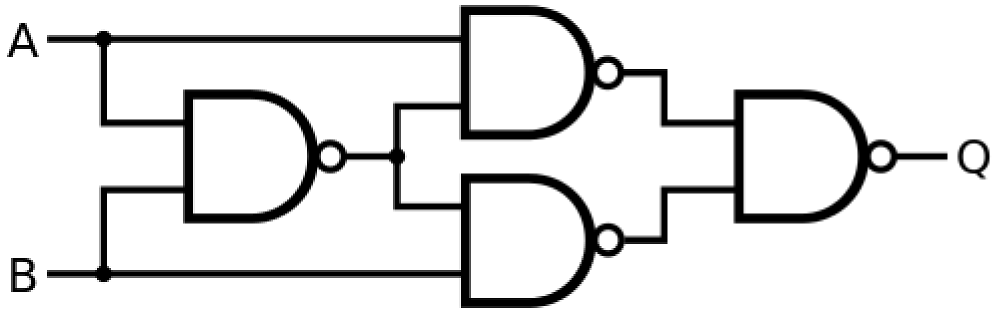


どのような考えで
分離するか

復習：NANDゲートで構成したXORゲート

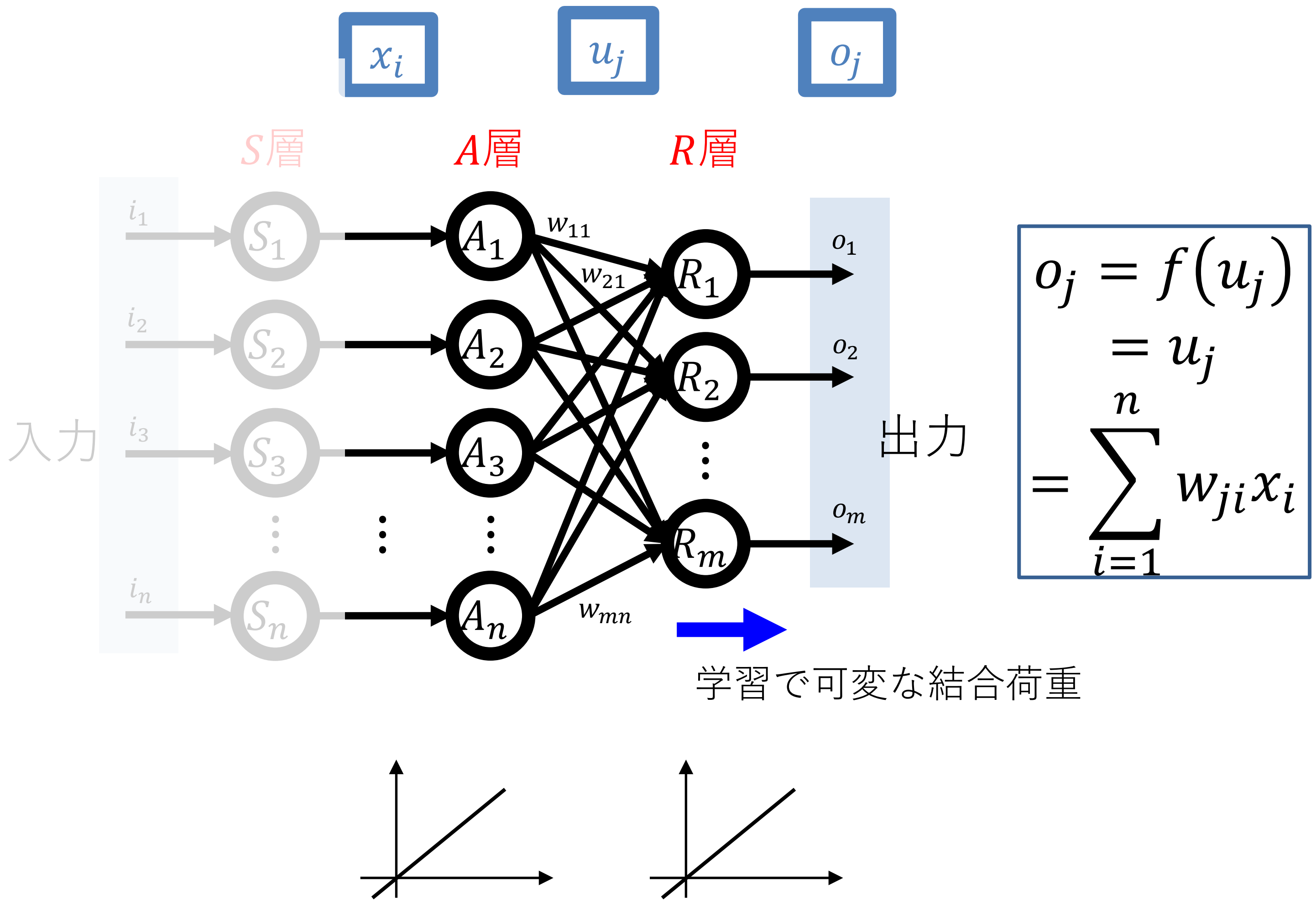
- XORゲートは線形分離可能なNANDゲートやNORゲートの組み合わせで構成可能

- 単純パーセプトロンにおいて線形分離不可能な問題は階層を深くすることにより分離可能となる



NANDゲートのみで構成したXORゲート

復習：線形ニューロンの2層ネットワーク



$f(u) = u$：線形の関数（何も変換しない）

復習：デルタ則　(Widrow-Hoff則)

別称：Δルール（教師あり学習）

連続値をとるA層とR層の2層からなるネットワークに、

誤り訂正学習法を拡張して適用

$$\boxed{\Delta w_{ji} = \eta \cdot (t_j - o_j) \cdot o_i}$$

η：学習率
（小さな正値）

w_{ji}：A層の第iユニットからR層間の第jユニットへの結合荷重

o_i：A層の第iユニットの出力値、連続値（$i = 1, \ldots, n$）

o_j：R層の第jユニットの出力値、連続値（$j = 1, \ldots, m$）

t_j：R層の第jユニットに対する教師信号、連続値

復習：デルタ則は最急降下法

最急降下法：関数$E(\mathbf{w})$の最小点（実際は極小点）を探す方法

$$\boxed{\Delta\mathbf{w} = -\eta \cdot \frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} \propto -\frac{\partial E(\mathbf{w})}{\partial \mathbf{w}}}$$

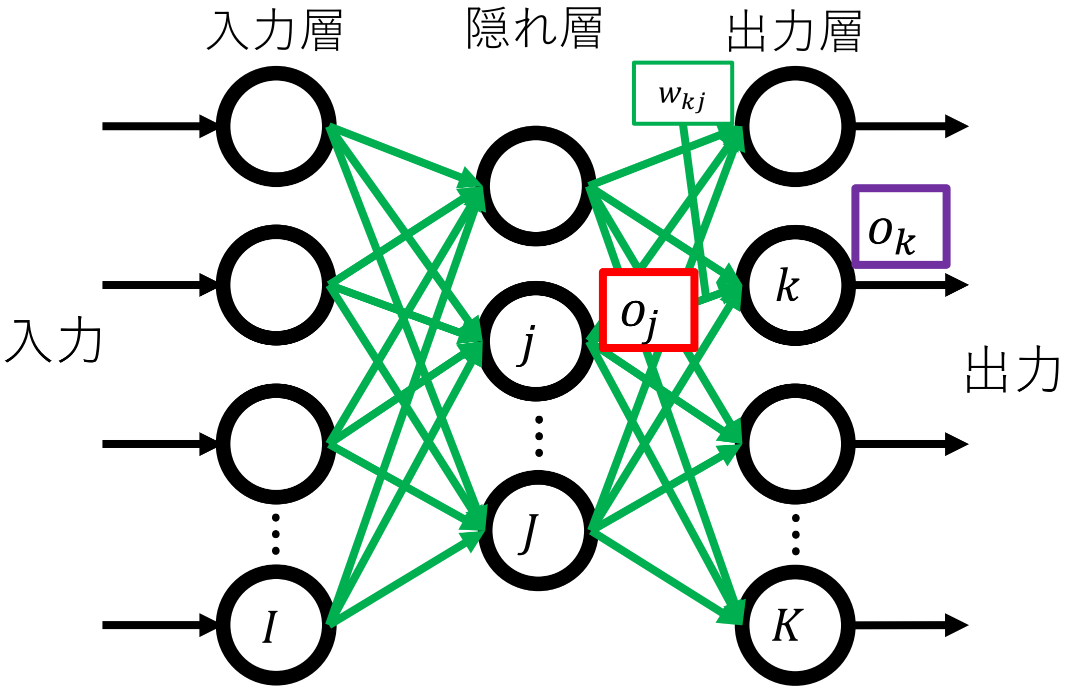
ここで、A層-R層の2層のネットワークに対する出力誤差の2乗和

誤差（の総和）

$$\boxed{E(\mathbf{w}) = \frac{1}{2}\sum_{j=1}^{m}\left(t_j - o_j\right)^2}$$

を考えると、デルタ則は最急降下法となっている

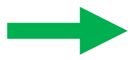
シグモイドニューロンの３層ネットワーク

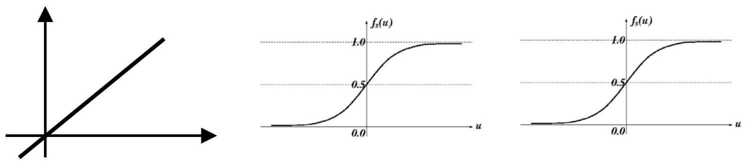


$$o_k = f(u_k)$$

$$u_k = \sum_{j=1}^{J} w_{kj} o_j$$

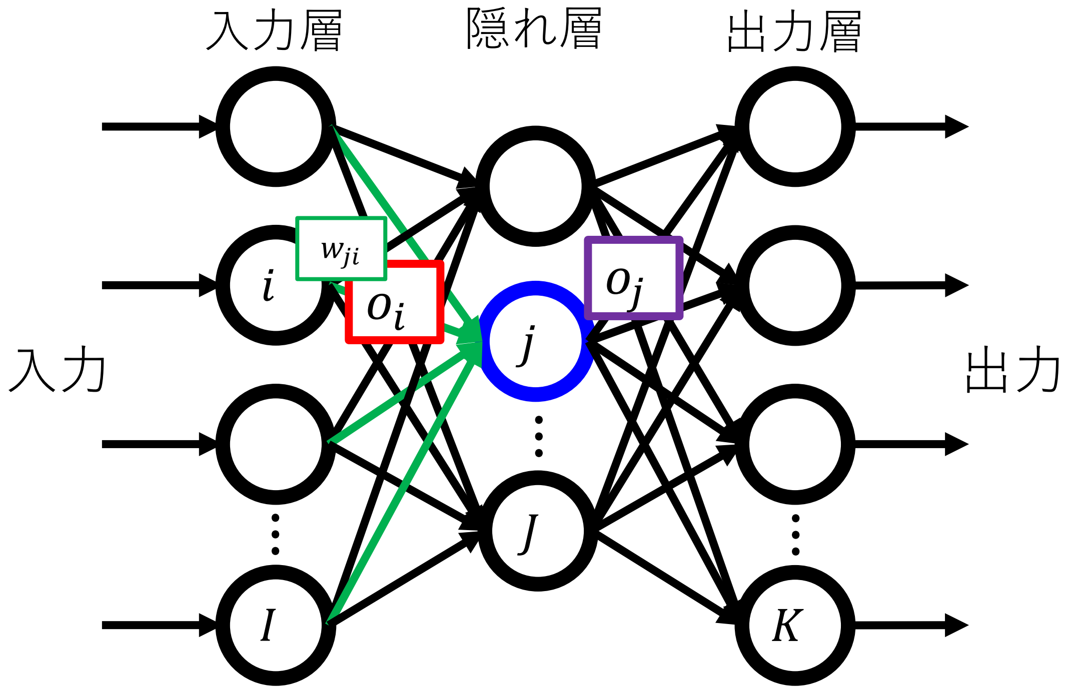
荷重和＋非線形変換





f：線形関数　f：シグモイド関数

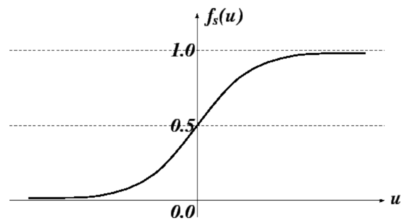
信号伝送過程（前進）
入力層ユニット → 隠れ層ユニット



$$o_j = f(u_j)$$

$$u_j = \sum_{i=1}^{I} w_{ji} o_i$$

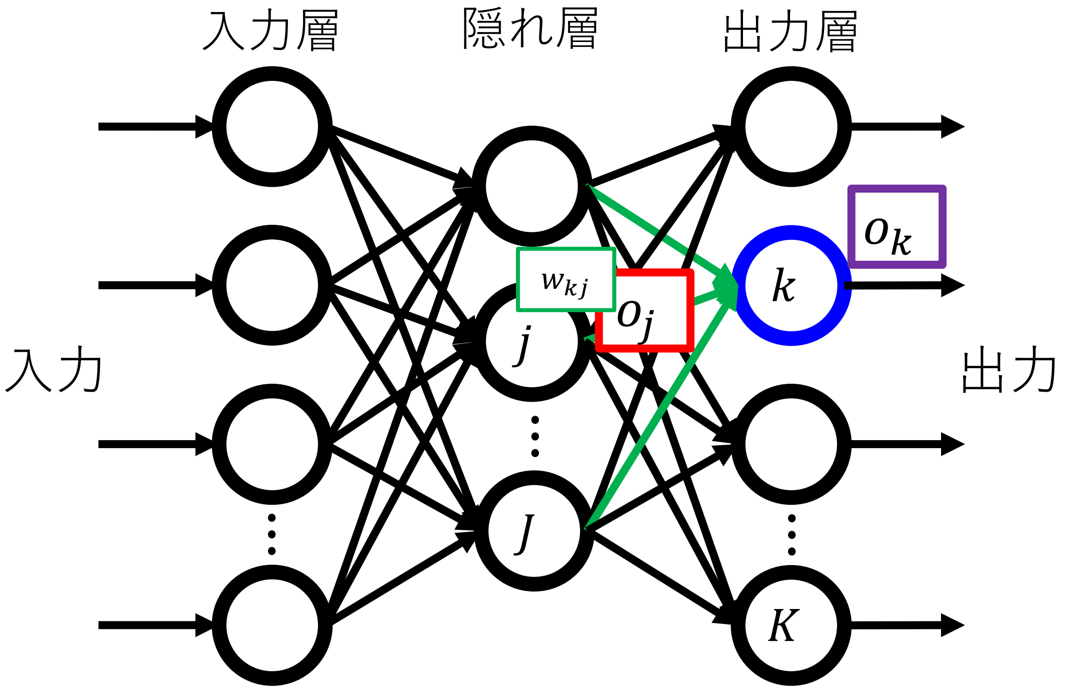
荷重和＋非線形変換

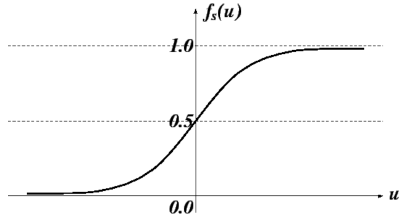


f：シグモイド関数

信号伝送過程（前進）

隠れ層ユニット → 出力層ユニット





f：シグモイド関数

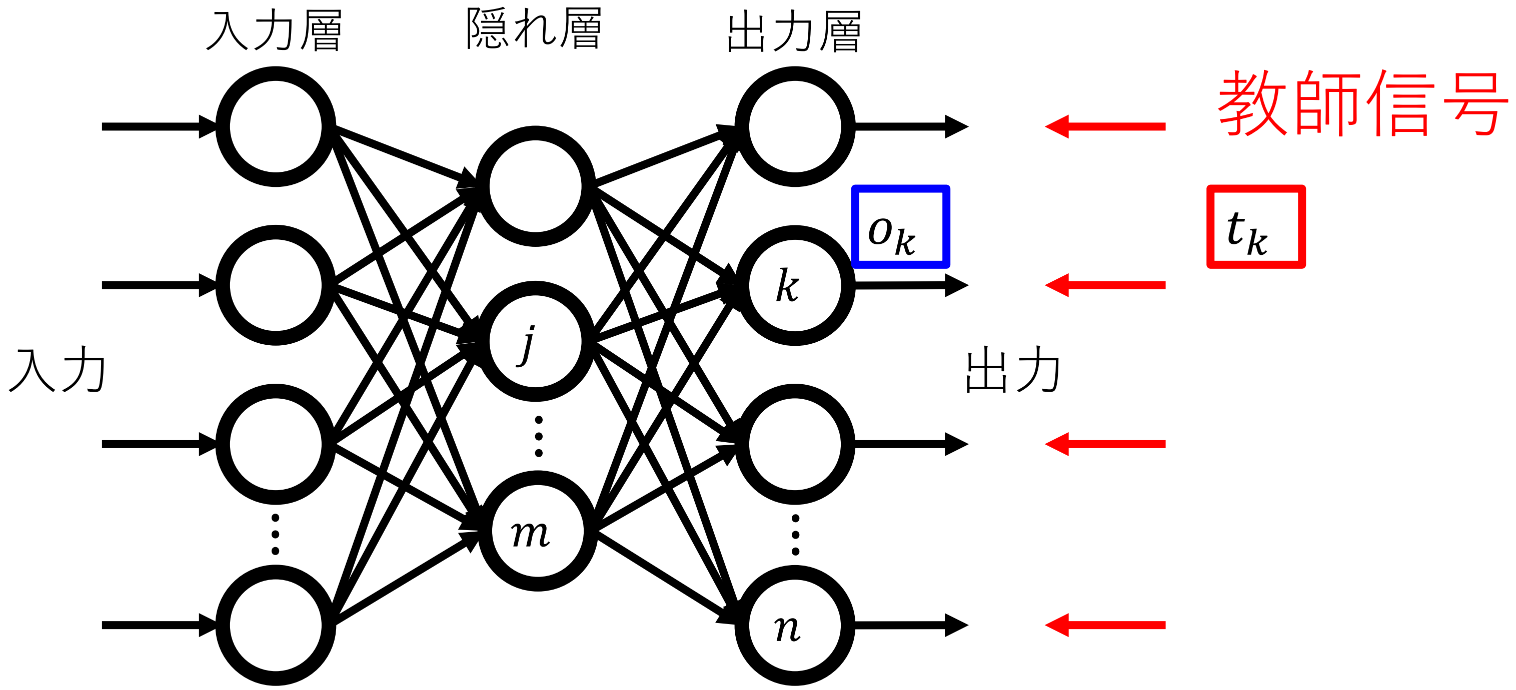
$$o_k = f(u_k)$$

$$u_k = \sum_{j=1}^{J} w_{kj} o_j$$

荷重和＋非線形変換

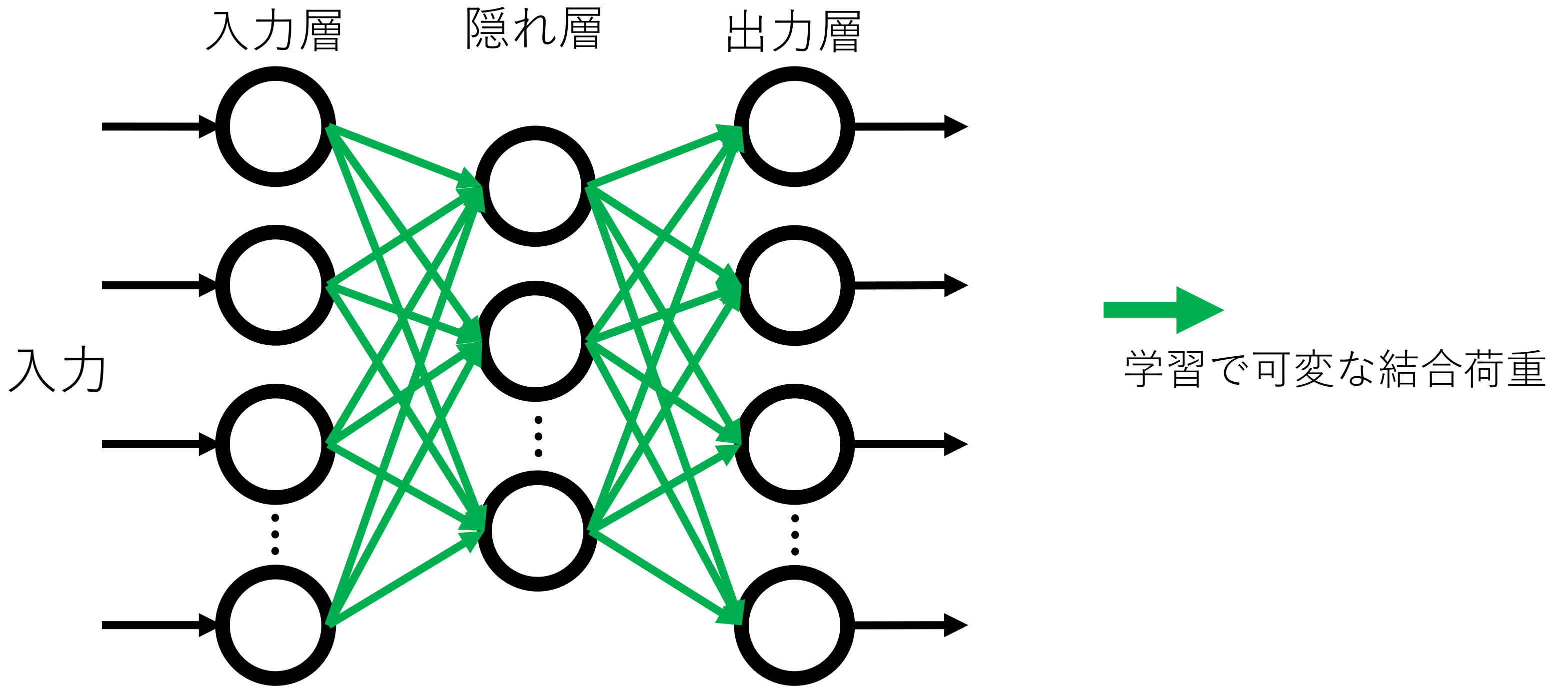
誤差の計算

出力値⇔教師信号 → 誤差$\boldsymbol{E}$



$$E(\mathbf{w}) = \frac{1}{2}\sum_{k=1}^{n}(t_k - o_k)^2$$

一般化デルタ則による誤差$\boldsymbol{E}$の最小化



最急降下法で $E(\mathbf{w})$ を小さくするように $\mathbf{w}$ を修正

$$\boxed{\text{修正後の}\mathbf{w} = \text{修正前の}\mathbf{w} + \text{修正量}\Delta\mathbf{w}}$$

一般化デルタ則も最急降下法

最急降下法：関数$E(\mathbf{w})$の最小点（実際は極小点）を探す方法

$$\boxed{\Delta\mathbf{w} = -\eta \cdot \frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} \propto -\frac{\partial E(\mathbf{w})}{\partial \mathbf{w}}}$$

出力誤差の2乗和

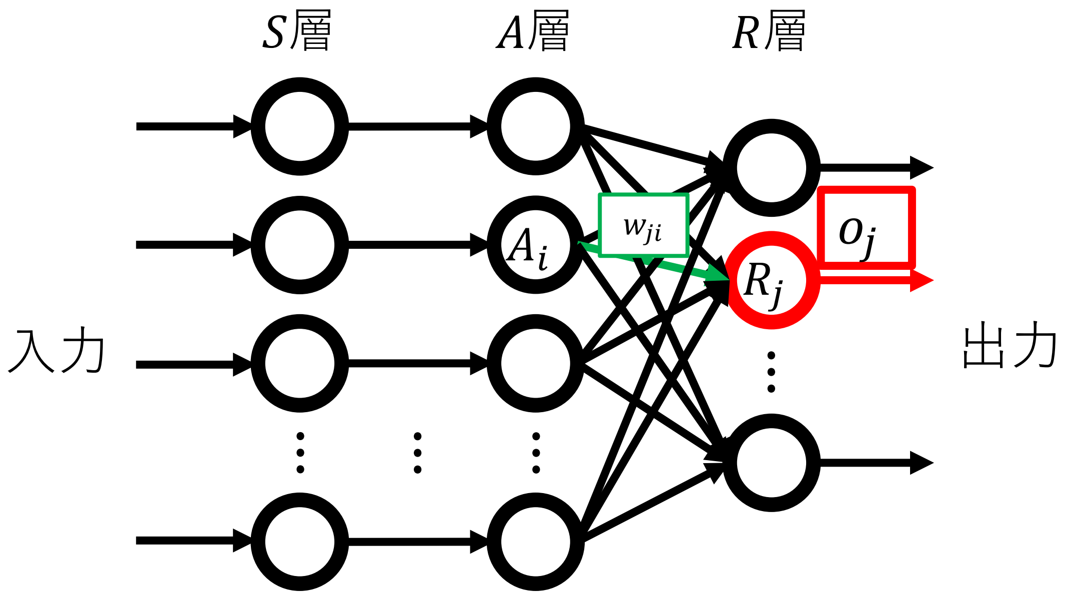
誤差（の総和）
$$\boxed{E(\mathbf{w}) = \frac{1}{2}\sum_{k=1}^{n}(t_k - o_k)^2}$$

を考え、最急降下法である一般化デルタ則を導く

復習：デルタ則

w_{ji}はo_jの誤差のみに責任があった

2層の場合は単純



復習：デルタ則

最急降下法

$$\Delta w_{ji} = -\eta \cdot \frac{\partial E(\mathbf{w})}{\partial w_{ji}}$$

$$= \eta \cdot \delta_j \cdot o_i$$

2層の場合は単純

誤差

$$E(\mathbf{w}) = \frac{1}{2}\sum_{j=1}^{n}\left(t_j - o_j\right)^2$$

最小化したい$\mathbf{w}$の関数E

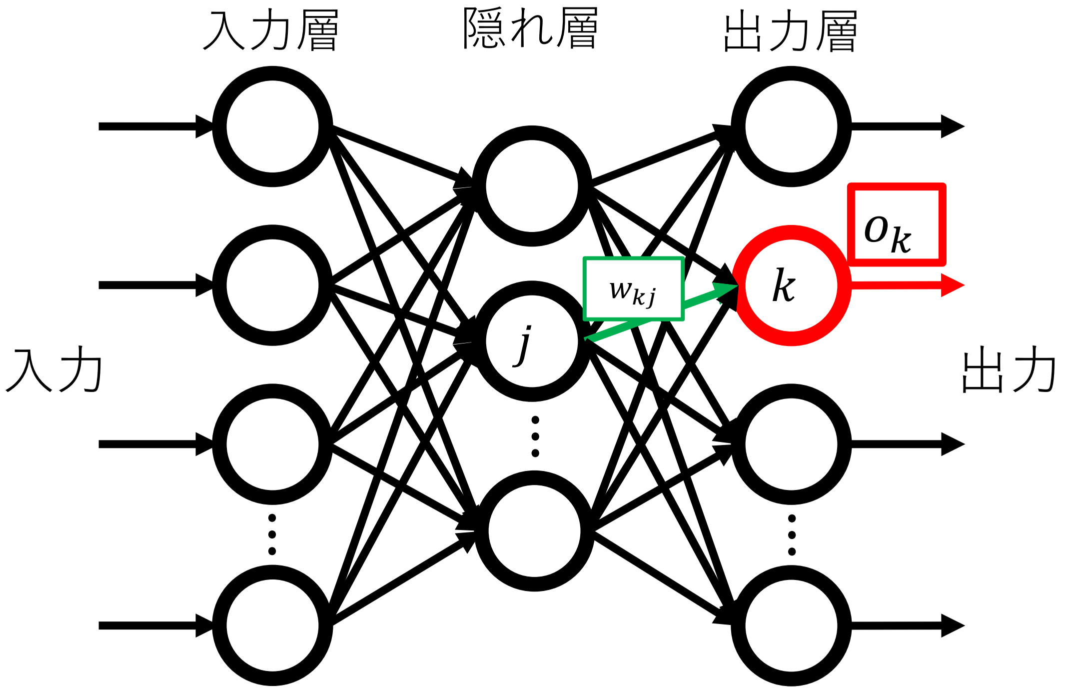
$$\begin{aligned}
\frac{\partial E(\mathbf{w})}{\partial w_{ji}} &= \frac{\partial}{\partial w_{ji}} \cdot \frac{1}{2}\sum_{j=1}^{n}\left(t_j - o_j(\mathbf{w})\right)^2 \\
&\rightarrow \frac{\partial}{\partial w_{ji}} \cdot \frac{1}{2}\left(t_j - o_j(w_{ji})\right)^2 \\
&= \left(t_j - o_j(w_{ji})\right) \cdot (-1) \cdot \frac{\partial o_j}{\partial w_{ji}} \\
&= -\left(t_j - o_j\right) \cdot o_i \\
&= -\delta_j \cdot o_i
\end{aligned}$$

合成関数の微分

$$o_j = \sum_{i=1}^{n} w_{ji} o_i$$

一般化デルタ則

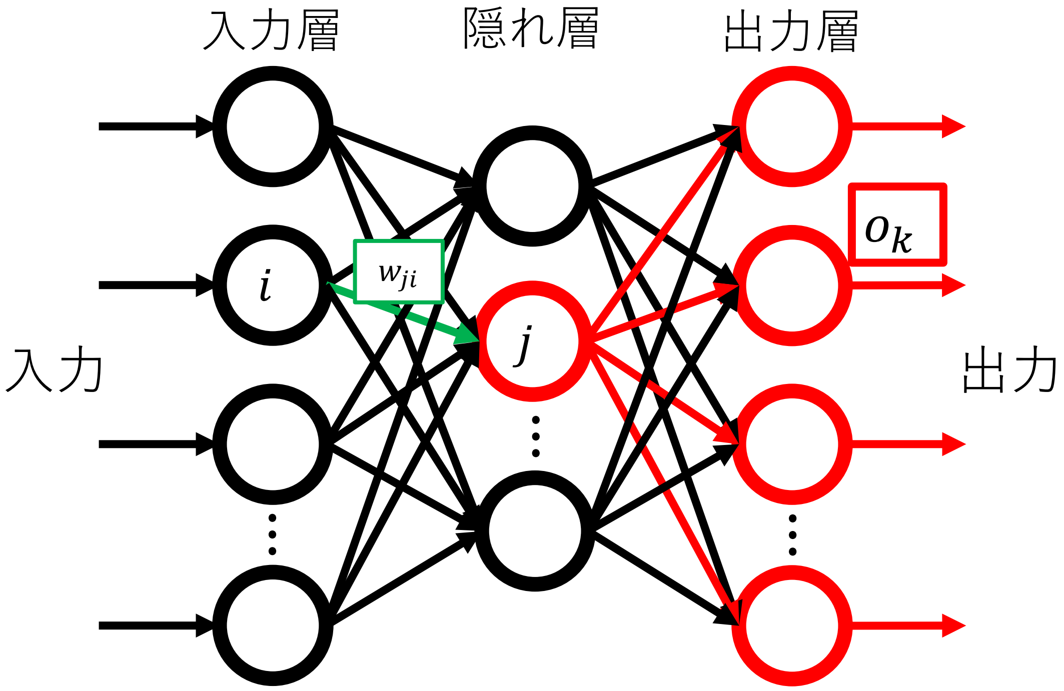
３層の場合、出力層のユニット$\boldsymbol{k}$の結合荷重w_{kj}は、
出力o_kと教師信号の誤差のみに影響し、責任がある



一般化デルタ則

3層の場合、隠れ層のユニット$\boldsymbol{j}$の結合荷重w_{ji}は全ての出力

o_k （$k = 1, \ldots, m$）に影響し、その誤差に責任がある



一般化デルタ則

入力層から隠れ層

$$o_j = f(u_j)$$

$$u_j = \sum_{i=1}^{n} w_{ji} o_i$$

隠れ層から出力層

$$o_k = f(u_k)$$

$$u_k = \sum_{j=1}^{m} w_{kj} o_j$$

$$o_k = f\left(\sum_{j=1}^{m} w_{kj} \cdot f\left(\sum_{i=1}^{n} w_{ji} o_i\right)\right)$$

合成関数の微分を順番に計算

一般化デルタ則

最急降下法

$$\Delta w_{ji} = -\eta \cdot \frac{\partial E(\mathbf{w})}{\partial w_{ji}}$$

$$= \eta \cdot \delta_j \cdot o_i$$

ここで、j は任意の層のj番目のユニット、i は一つ前の層のi番目のユニットとする。

誤差

$$E(\mathbf{w}) = \frac{1}{2}\sum_{j=1}^{J}(t_k - o_k)^2$$

最小化したい$\mathbf{w}$の関数E

$$\frac{\partial E(\mathbf{w})}{\partial w_{ji}} = \frac{\partial}{\partial w_{ji}} \cdot \frac{1}{2}\sum_{j=1}^{J}\left(t_k - o_k(\mathbf{w})\right)^2$$

$$\rightarrow \frac{\partial}{\partial w_{ji}} \cdot \frac{1}{2}(t_k - f(u_k))^2$$

$=???$

$???$

$-\delta_j \cdot o_i$

合成関数の微分

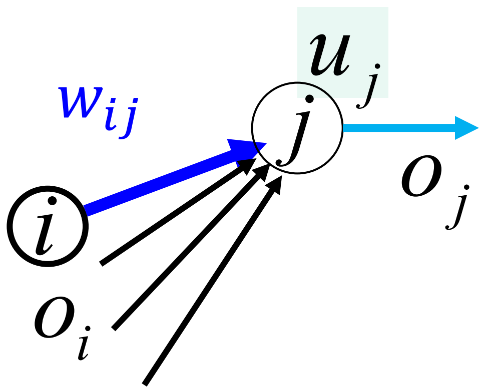
$$\frac{\partial E(\mathbf{w})}{\partial w_{ji}} = \frac{\partial E(\mathbf{w})}{\partial u_j} \cdot \frac{\partial u_j}{\partial w_{ji}}$$

$$= \frac{\partial E(\mathbf{w})}{\partial u_j} \cdot \frac{\partial \left(\sum_i w_{ji} o_i \right)}{\partial w_{ji}}$$

$$= \frac{\partial E(\mathbf{w})}{\partial u_j} \cdot o_i$$

$$= -\delta_j \cdot o_i$$

$$u_j = \sum_{i=1} \boxed{w_{ji}} o_i$$



$$\delta_j \equiv -\frac{\partial E}{\partial u_j}$$

全体の誤差δのうち、u_jに関するものを抜き出し

したがって

$$\frac{\partial E(\mathbf{w})}{\partial w_{ji}} = -\delta_j \cdot o_i$$

最急降下法を用いて以下を解く

$$w'_{ji} = w_{ji} - \eta \cdot \delta_j \cdot o_i$$

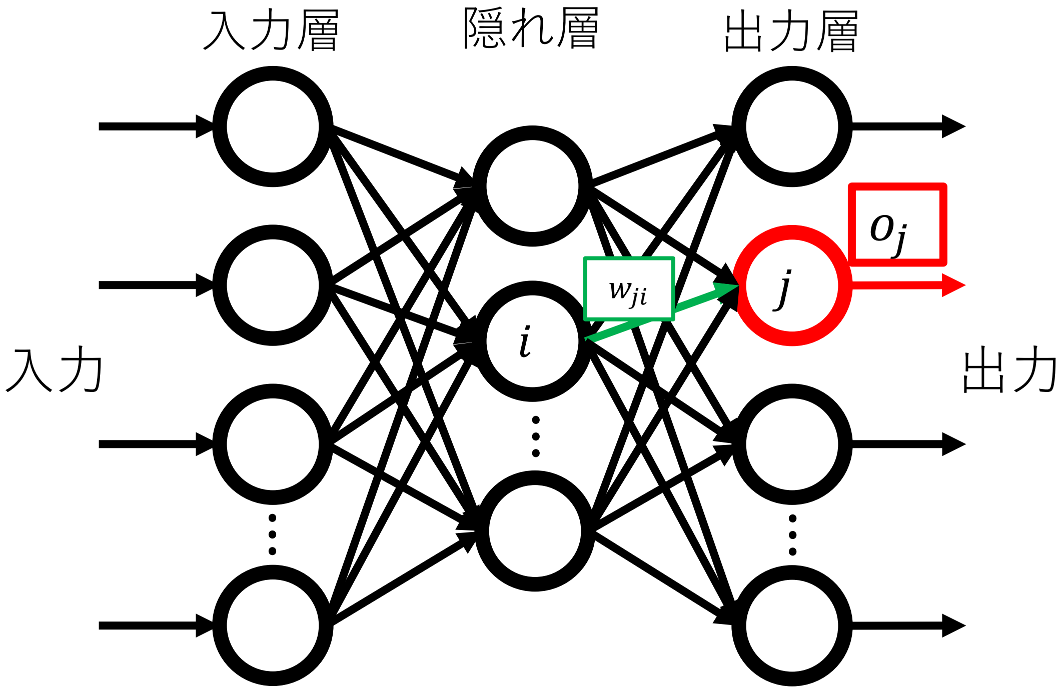
各ユニットに対して、δ_jを決定する

$$\begin{aligned}
\delta_j &\equiv -\frac{\partial E}{\partial u_j} \\
&= -\frac{\partial E}{\partial o_j} \cdot \frac{\partial o_j}{\partial u_j} \\
&= -\frac{\partial E}{\partial o_j} \cdot \frac{\partial f(u_j)}{\partial u_j} \\
&= -\boxed{\frac{\partial E}{\partial o_j}} \cdot f'(u_j)
\end{aligned}$$

$$\boxed{o_j = f(u_j)}$$

一般化デルタ則

3層でも、出力層のユニットjの結合荷重w_{ji}は、出力o_jの誤差のみに影響し、責任がある



j を出力層ユニットとしたとき

$$\boxed{\frac{\partial E}{\partial o_j} = \frac{\partial}{\partial o_j} \cdot \frac{1}{2} \sum_{k=1}^{m} \left(t_k - o_k\right)^2}$$

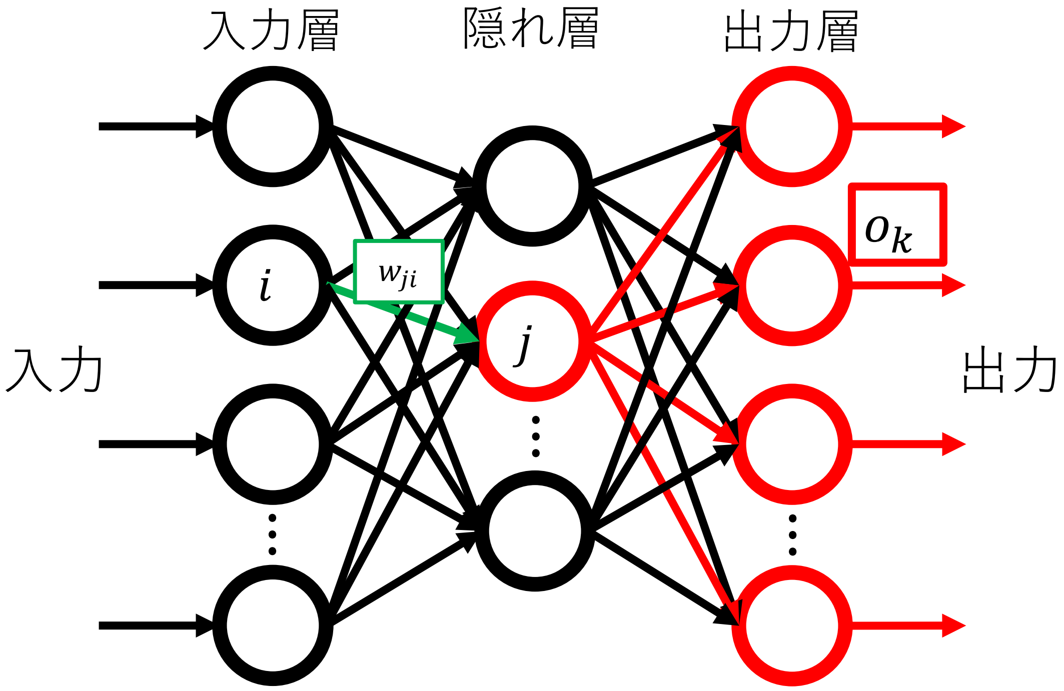
$$\frac{\partial E}{\partial o_j} = \frac{\partial}{\partial o_j} \cdot \frac{1}{2} \left(t_j - o_j\right)^2$$

$$= \frac{1}{2} \cdot 2 \cdot (-1) \cdot \left(t_j - o_j\right) = -\left(t_j - o_j\right)$$

$$\therefore \delta_j = \left(t_j - o_j\right) \cdot f'\left(u_j\right)$$

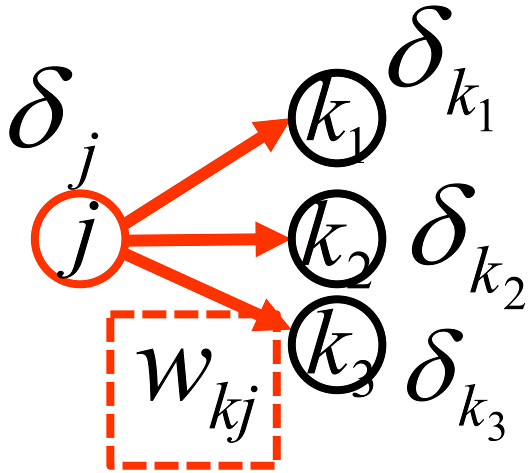
一般化デルタ則

3層の場合、隠れ層のユニットjもあり、結合荷重w_{ji}は全ての出力o_k（$k = 1, \ldots, m$）に影響し、その誤差に責任がある



j が隠れ層ユニットのとき

j が出力を送っている全てのm



$$\frac{\partial E}{\partial o_j} = \sum_k \frac{\partial E}{\partial u_k} \cdot \frac{\partial u_k}{\partial o_j}$$

$$= \sum_k \frac{\partial E}{\partial u_k} \cdot \frac{\partial}{\partial o_j} \sum_j w_{kj} o_j$$

$$= \sum_k \frac{\partial E}{\partial u_k} \cdot w_{kj}$$

$$= \sum_k (-\delta_k) \cdot w_{kj}$$

$$\therefore \delta_j = f'(u_j) \cdot \sum_k \delta_k \cdot \boxed{w_{kj}}$$

- j を出力層ユニットとしたとき

$$\boxed{\delta_j = (t_j - o_j) \cdot f'(u_j)}$$

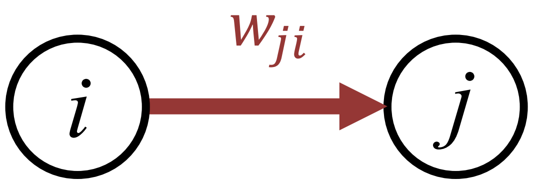
- j が隠れ層ユニットのとき

$$\boxed{\delta_j = f'(u_j) \cdot \sum_k \delta_k \cdot w_{kj}}$$

一般化デルタ則

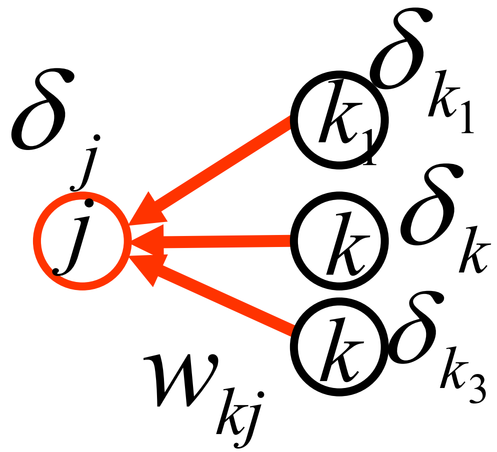
$$\boxed{\Delta w_{ji} = \eta \cdot \delta_j \cdot o_i}$$

送り手　　受け手



出力値　　誤差

o_i　　δ_j



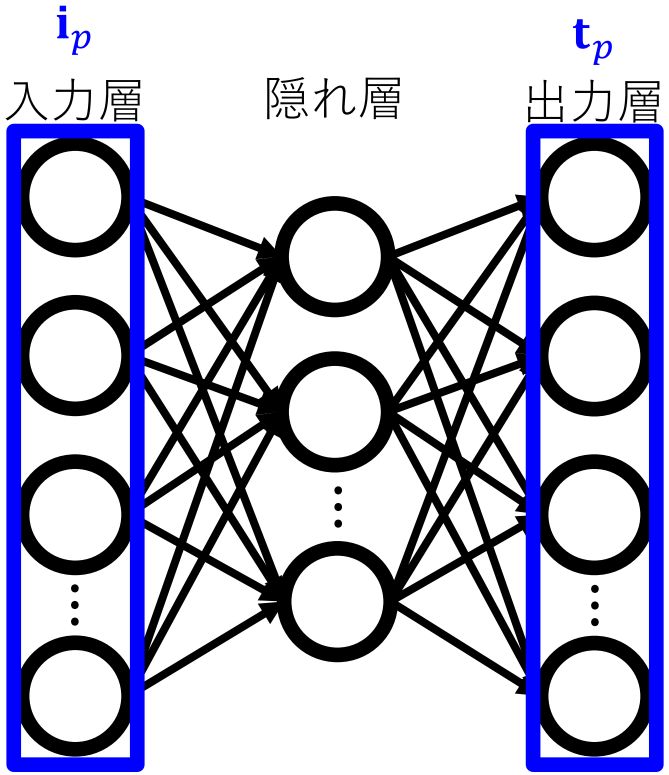
誤差は

$$\boxed{\delta_j = f'(u_j) \cdot \sum_k \delta_k \cdot w_{kj}}$$

逆方向（出力層から入力層へ）に伝えられてゆく：学習過程（後進）

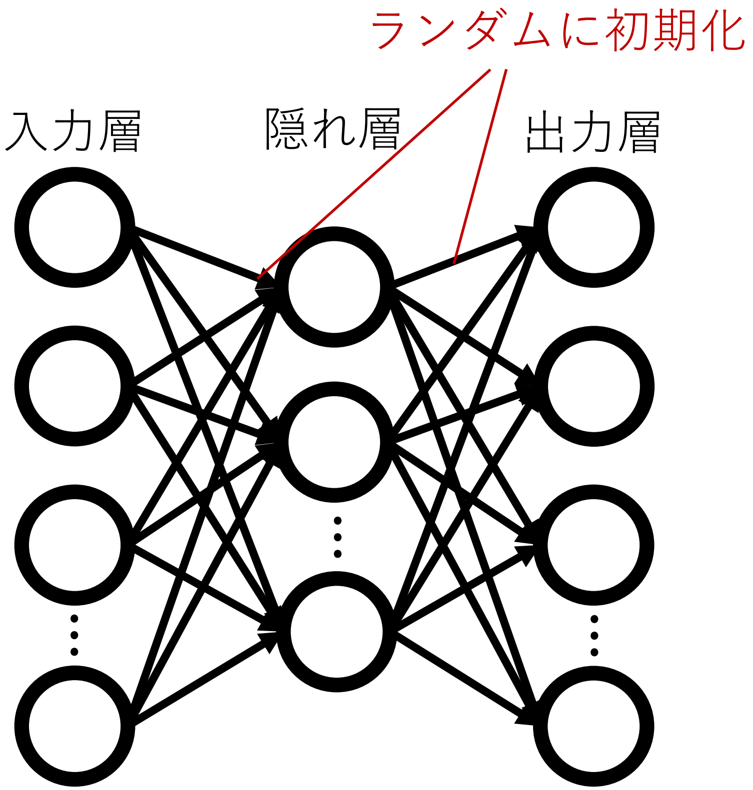
誤差逆伝搬法（Back Propagation法）(1/6)

1. 学習を行うための入力パターンと対応する目標出力の組の集合を考える



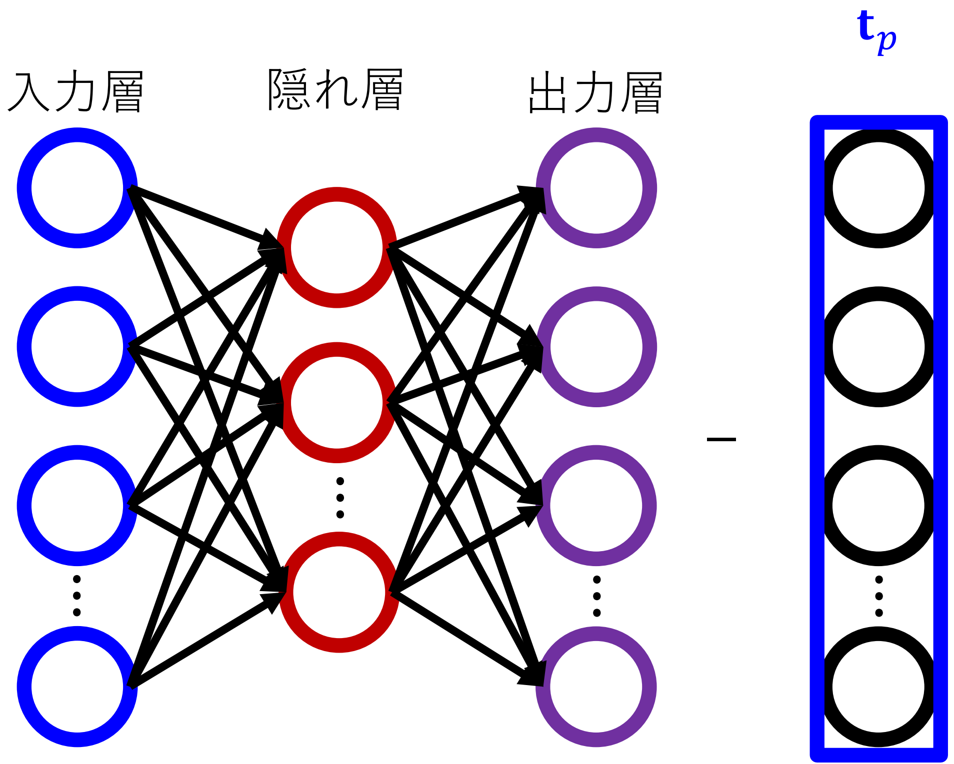
誤差逆伝搬法（Back Propagation法）（2/6）

2. 入力層、中間層、出力層の順に、各ユニットの入出力を計算する
   この操作は信号伝送過程、あるいは、誤差逆伝搬法における
   前進型処理とよばれる



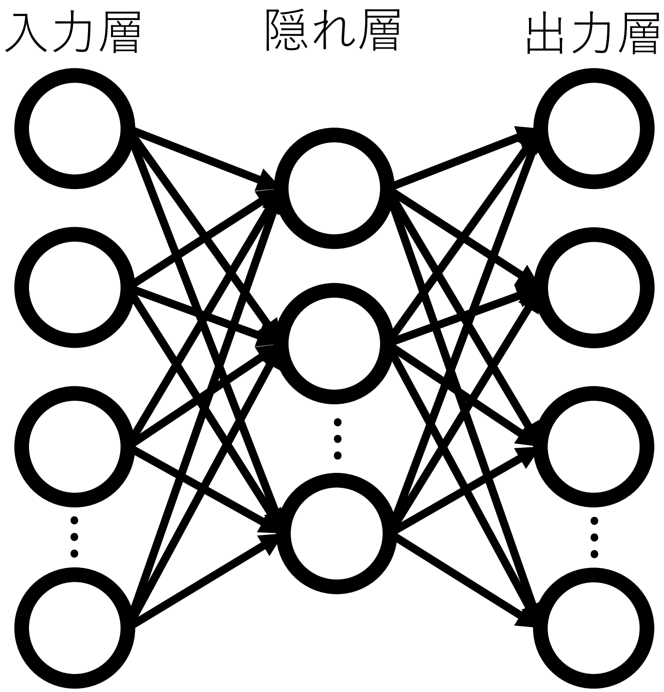
誤差逆伝搬法（Back Propagation法）(3/6)

3. 入力パターンに対する目標出力である教師信号と、実際に得られた出力との2乗誤差を計算する



誤差逆伝搬法（Back Propagation法）(4/6)

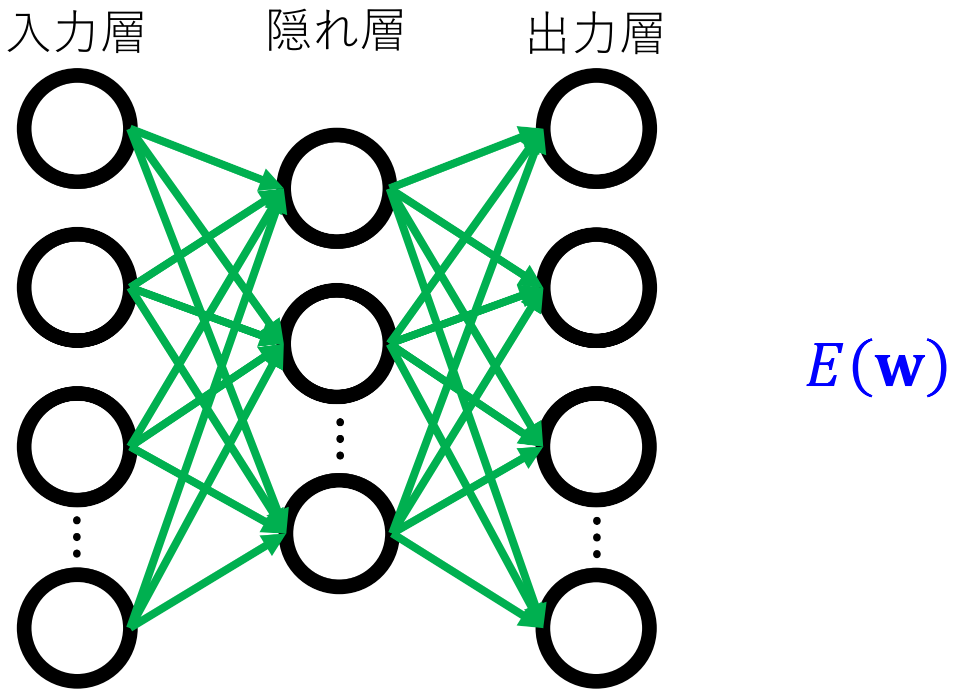
4. 得られた誤差を最小化するように、出力層から入力層に向かって、ネットワークの各層間の結合荷重を修正する
この操作は学習過程、あるいは、誤差逆伝搬法における後進型処理とよばれる



$E(\mathbf{w})$

誤差逆伝搬法（Back Propagation法）（5/6）

5. すべての入力パターンに対する2乗誤差が設定値以下になれば学習が収束したと判断して終了する
   そうでなければ、手順2から手順4の操作を繰り返す



誤差逆伝搬法（Back Propagation法）（6/6）

入力層から隠れ層1

$$o_j = f(u_j)$$

$$u_j = \sum_{i=1}^{n'} w_{ji} o_i$$

隠れ層が
何層あっても

隠れ層zから出力層

$$o_k = f(u_k)$$

$$u_k = \sum_{j=1}^{n} w_{kj} o_j$$

$$o_k = f\left(\sum_{j=1}^{n} w_{kj} \cdot f\left(\sum \cdots \sum w \cdot f\left(\sum_{i=1}^{n'} w_{ji} o_i\right) \cdots\right)\right)$$

要は、合成関数の微分を順番に計算するだけ

まとめ

- シグモイドニューロンの三層ネットワークについて説明した。
- 隠れ層と出力層におけるパラメータを学習する一般化デルタ則について説明した。
- 一般化デルタ則に基づいて多層ニューラルネットワークの学習を可能とする誤差逆伝搬法について説明した。

復習問題

1. Ｓ字型の活性化関数を何というか？

2. 多層のニューロンの結合荷重の修正量を計算するアルゴリズムは何か？

3. 上記アルゴリズムに基づいて出力層から入力層に向かって各層の結合荷重を修正する操作（学習過程）を何というか？

4. 上記の学習過程は、どういった関数を微分する事で達成されるか？

次回の講義

- 復習問題と解説
  - 前半部分の問題を復習する