

バックプロパゲーション（誤差逆伝播法）

(有)数理解析研究所 一色 浩

1. XOR問題

図1のような単純パーセプトロンでは、排他的論理和（XOR）問題は絶対に解けない．排他的論理和とは、2 つの入力が（1, 0）または（0, 1）のときのみ1を出力する問題である。図2左を見るとわかるとおり、XOR 問題は一本の判別直線で白マルと黒マルを分離できない、すなわち線形分離不可能な問題である。

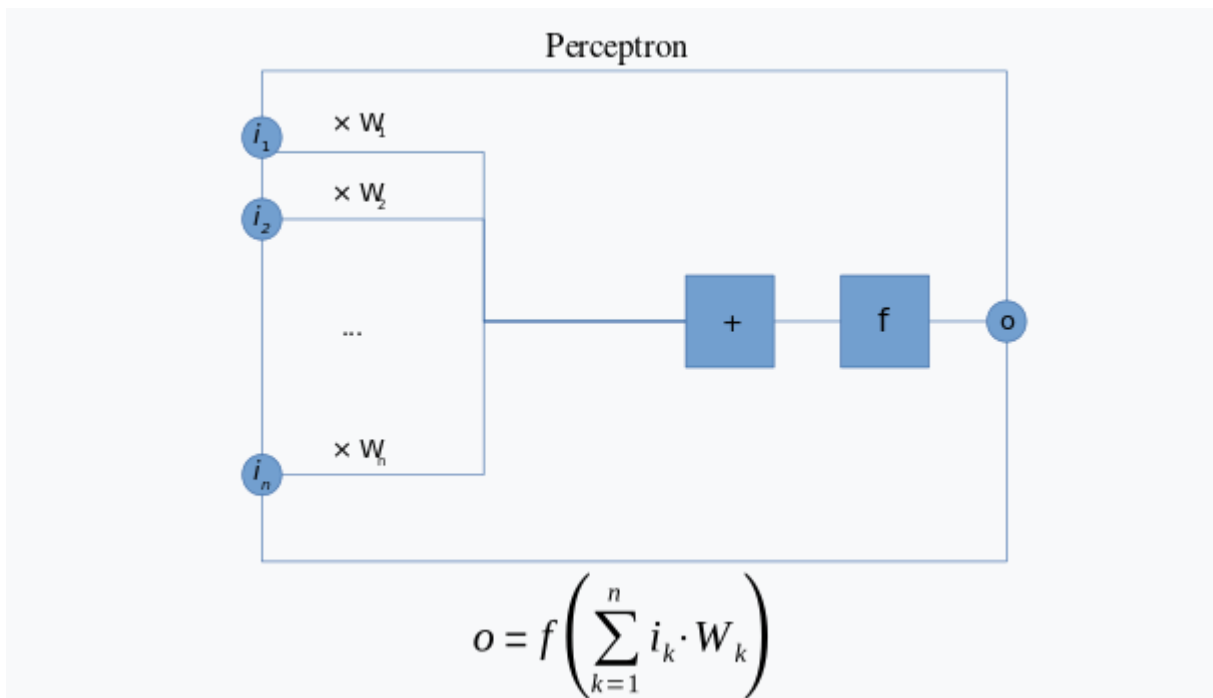


図 1 単純パーセプトロン (<https://en.wikipedia.org/wiki/Perceptron>)

図 1 に示されるように、単純パーセプトロンで示されるのは、1 次式であり 2 次元の場合は一本の直線である．この直線の上か下かで出力の値が異なる．これを点の判別に利用できるが、XOR の場合には直線が 2 本いることになり、一本の直線しか使えない単純パーセプトロンでは、XOR の判別はできないことになる．

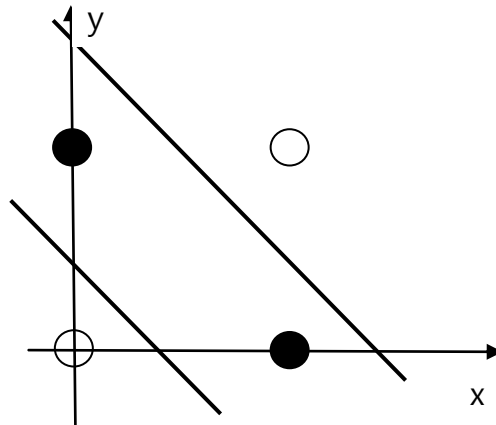


図2 線形分離不可能な問題

2. バックプロパゲーション

m 層からなるニューラルネットワークを考える．第 m 層のニューロンの数を N_m とする．図1にニューロンのネットワークを示す．

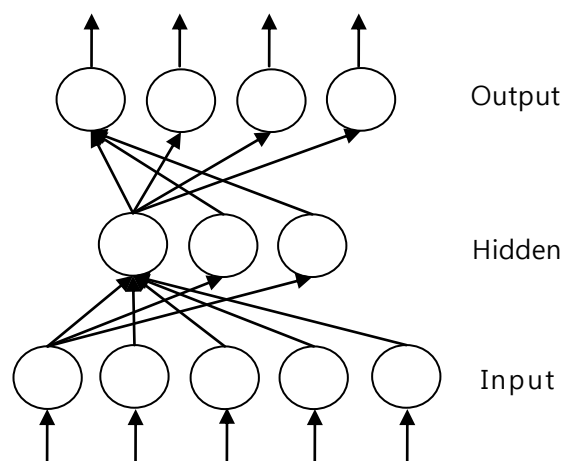


図1 ニューラルネットワーク（煩雑になるので，層の上下のニューロンは通常すべて結線される．また，簡単のためにバイアスを与えるニューロンも省略してある）

図2に一つのニューロンの入出力を示す．第 k 層のニューロンの数を N_k とし，第 k 層の番号 j のニューロン C_j^k の第 $k-1$ 層の番号 i の出力 y_i^{k-1} からの入力を x_{ij}^k ，第 k 層のニューロンの出力を y_j^k とする． x_{ij}^k をすべて足し合わせたものが，

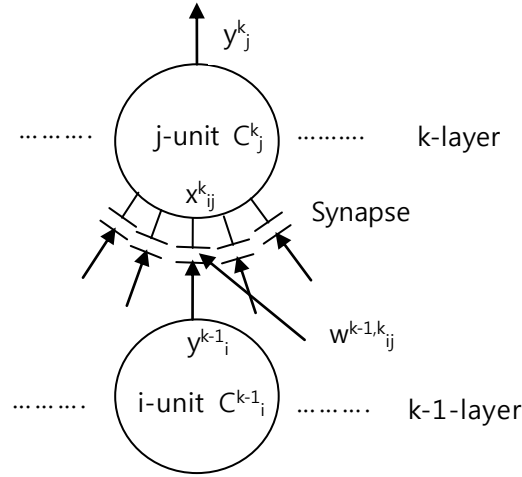


図2 ニューロンの入出力と変数

$$y_j^k = f(x_j^k) = \frac{1}{1 + e^{-x_j^k}} \quad (1)$$

$$x_j^k = \sum_i^{N_{k-1}} x_{ij}^k = \sum_i^{N_{k-1}} w_{ij}^{k-1,k} y_i^{k-1} \quad (2)$$

ここで、(1)式の最右辺の関数はシグモイド関数と呼ばれる非線形関数である．第 k 層の第 j 番目のニューロン C_j^k の入力と第 $k-1$ 層の第 i 番目のニューロン C_i^{k-1} の出力 y_i^{k-1} を結ぶ軸索のウェイトを $w_{ij}^{k-1,k}$ としている．添字の変域を以下に示す．

$$k = 2, \dots, N_k, \quad i = 1, 2, \dots, N_{k-1}, \quad j = 1, 2, \dots, N_k \quad (\text{ISK1})$$

出力層である第 m 層の番号 j のニューロン C_j^m に与えられる教師データを t_j とすると，誤差 δ_j^m とコスト関数 E は下式で定義される．

$$\delta_j^m = y_j^m - t_j \quad (3a)$$

$$E = \frac{1}{2} \sum_j (\delta_j^m)^2 \quad (3b)$$

コスト関数 E は，第 k 層の第 j 番目のニューロン C_j^k の入力と第 $k-1$ 層の第 i 番目のニューロン C_i^{k-1} の出力を結ぶ軸索のウェイトを $w_{ij}^{k-1,k}$ とし，コスト関数 E を最小にする最急降下法の算式は ε をパラメーターとして

$$\Delta w_{ij}^{k-1,k} = -\varepsilon \frac{\partial E}{\partial w_{ij}^{k-1,k}} \quad (4)$$

により与えられる．ウェイト $w_{ij}^{k-1,k}$ の更新式は

$$\left[w_{ij}^{k-1,k} \right]_{\text{new}} = \left[w_{ij}^{k-1,k} \right]_{\text{old}} + dw_{ij}^{k-1,k} \quad (\text{ISK2})$$

となる．

(4)式の微分をチェインルールを使って書き直すと

$$\frac{\partial E}{\partial w_{ij}^{m-1,m}} = \sum_{l=1}^{N_k} \frac{\partial E}{\partial y_l^m} \frac{\partial y_l^m}{\partial w_{ij}^{m-1,m}} = \frac{\partial E}{\partial y_j^m} \frac{\partial y_j^m}{\partial w_{ij}^{m-1,m}} = \delta_j^m \frac{\partial y_j^m}{\partial w_{ij}^{m-1,m}} \quad (*5)$$

のように書き直せる．まず次の微分を考える． y_j^m は x_j^m のみの関数であるから

$$\frac{\partial y_j^m}{\partial w_{ij}^{m-1,m}} = \frac{\partial y_j^m}{\partial x_j^m} \frac{\partial x_j^m}{\partial w_{ij}^{m-1,m}} \quad (6)$$

ここで，出力関数がシグモイド関数の場合には，すなわち， $y = 1/(1+e^{-x})$ のとき

$$\frac{\partial y}{\partial x} = \frac{\partial}{\partial x} \frac{1}{1+e^{-x}} = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{1+e^{-x}} \left(1 - \frac{1}{1+e^{-x}} \right) = y(1-y) \quad (7)$$

に注意しておく．

つぎに， $\Delta w_{ij}^{m-1,m}$ を書き直す．

$$\Delta w_{ij}^{m-1,m} = -\varepsilon \frac{\partial E}{\partial w_{ij}^{m-1,m}} \quad (8)$$

$$= -\varepsilon \sum_{l=1}^{N_k} \frac{\partial E}{\partial y_l^m} \frac{\partial y_l^m}{\partial w_{ij}^{m-1,m}} = -\varepsilon \frac{\partial E}{\partial y_j^m} \frac{\partial y_j^m}{\partial w_{ij}^{m-1,m}} = -\varepsilon \frac{\partial E}{\partial y_j^m} \frac{\partial y_j^m}{\partial x_j^m} \frac{\partial x_j^m}{\partial w_{ij}^{m-1,m}} \quad (9)$$

$$\begin{aligned} &= -\varepsilon \frac{\partial E}{\partial y_j^m} y_j^m (1-y_j^m) \frac{\partial \sum_{k=1}^{N_{m-1}} w_{kj}^{m-1,m} y_k^{m-1}}{\partial w_{ij}^{m-1,m}} \\ &= -\varepsilon \frac{\partial E}{\partial y_j^m} y_j^m (1-y_j^m) \sum_{k=1}^{m-1} d_k^i y_k^{m-1} = -\varepsilon (y_j^m - t_j) y_j^m (1-y_j^m) y_i^{m-1} \end{aligned} \quad (*10)$$

$$= -\varepsilon \delta_j^m y_j^m (1 - y_j^m) y_i^{m-1} \quad (*11)$$

ここで、 d_j^i はChronockerのデルタで

$$d_j^i = \begin{cases} 1 & \text{when } i = j \\ 0 & \text{otherwise} \end{cases} \quad (\text{ISK3})$$

である。

つぎに

$$\delta_j^n = \frac{\partial E}{\partial y_j^n} \quad (\text{ISK4})$$

と書いて、第 n 層の誤差と考える。すると

$$\delta_j^n = \frac{\partial E}{\partial y_j^n} = \sum_{k=1}^{N_{n+1}} \frac{\partial E}{\partial y_k^{n+1}} \frac{\partial y_k^{n+1}}{\partial y_j^n} = \sum_{k=1}^{N_{n+1}} \frac{\partial E}{\partial y_k^{n+1}} \frac{\partial y_k^{n+1}}{\partial x_k^{n+1}} \frac{\partial x_k^{n+1}}{\partial y_j^n} \quad (12)$$

$$= \sum_{k=1}^{N_{n+1}} \frac{\partial E}{\partial y_k^{n+1}} \frac{\partial y_k^{n+1}}{\partial x_k^{n+1}} w_{jk}^{n,n+1} = \sum_{k=1}^{N_{n+1}} \frac{\partial E}{\partial y_k^{n+1}} y_k^{n+1} (1 - y_k^{n+1}) w_{jk}^{n,n+1} \quad (*13)$$

$$= \sum_{k=1}^{N_{n+1}} \delta_k^{n+1} y_k^{n+1} (1 - y_k^{n+1}) w_{jk}^{n,n+1} \quad (14)$$

が得られる。ウェイトの修正量 $\Delta w_{ij}^{k-1,k}$ については、(11)式と同様にして

$$\Delta w_{ij}^{k-1,k} = -\varepsilon \frac{\partial E}{\partial w_{ij}^{k-1,k}} = -\varepsilon \delta_j^k y_j^k (1 - y_j^k) y_i^{k-1} \quad (15)$$

となる。

以上をまとめると、(3a)式、(14)、(15)式から

$$\delta_j^k = \begin{cases} y_j^k - t_j & \text{if } k = m \\ \sum_{l=1}^{N_{k+1}} \delta_l^{k+1} y_l^{k+1} (1 - y_l^{k+1}) w_{jl}^{k,k+1} & \text{otherwise} \end{cases} \quad (*16)$$

$$\Delta w_{ij}^{k-1,k} = -\varepsilon \delta_j^k y_j^k (1 - y_j^k) y_i^{k-1} \quad (\text{ISK5})$$

というアルゴリズムが得られる。

(2)式と(16)式を比べると、信号 y_j^k の伝播は入力層から出力層に、すなわち

$k=1 \rightarrow m$ にかかる．一方， δ_j^k 誤差はこれと逆方向にかかることが分かる．これが誤差逆伝播と呼ばれる所以である．

3. MatLabによるコード

作成中

4. 計算例

作成中

5. 結言

作成中

参考文献

[1] 浅川伸一，バックプロパゲーション，pp.1-20，

<http://www.cis.twcu.ac.jp/~asakawa/waseda2002/bp.pdf>