第７章　パーセプトロン型学習規則

本：はじめてのパターン認識
著：平井　有三

茨城大学大学院理工学研究科情報工学専攻
菊池　裕紀

導入

- パーセプトロンの学習規則
  - 2クラスの線形識別関数を求める古典的な方法
  - パーセプトロンの収束定理を利用
    - 識別関数のパラメータを逐次的に求めるアルゴリズムでは、2クラスの学習データが線形分離可能であれば収束

パーセプトロン(1)

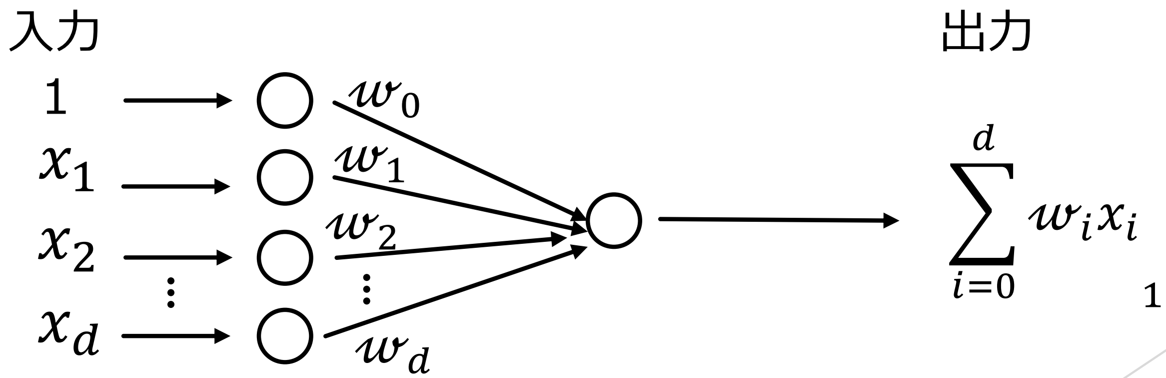
▶ パーセプトロンの学習規則

線形分離関数 ： $f(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x}$

$$\boldsymbol{w} = (w_0, w_1, \dots, w_d)^T,\ \boldsymbol{x} = (x_0 = 1, x_1, \dots, x_d)^T$$

$f(\boldsymbol{x}) \geq 0$のとき$\boldsymbol{x} \in C_1$

$f(\boldsymbol{x}) < 0$のとき$\boldsymbol{x} \in C_2$



1

パーセプトロン(2)

- 係数ベクトルwの調整

$f(\boldsymbol{x}_i) \geq 0$であれば　$\boldsymbol{w}_{i+1} = \boldsymbol{w}_i$

$f(\boldsymbol{x}_i) < 0$であれば　$\boldsymbol{w}_{i+1} = \boldsymbol{w}_i + \eta \boldsymbol{x}_i$

η：学習の収束速度を決めるパラメータ

η=1のとき、固定増分誤り訂正

パーセプトロン(3)

- 学習の難しさの尺度
  - ノイズがある場合、誤差が生じやすくなる
  - 識別超平面からある値$h > 0$より近い距離にあれば誤り
    - wを更新する
    - この h をマージンという

パーセプトロン(4)

- 係数ベクトルw_iの変化量Δw_i

$$\Delta w_i = \eta f_{step}\left(h - \frac{\boldsymbol{w}_i \boldsymbol{x}_i}{||\boldsymbol{w}_i||}\right)\boldsymbol{x}_i = \begin{cases} \eta \boldsymbol{x}_i & \left(h > \frac{\boldsymbol{w}_i \boldsymbol{x}_i}{||\boldsymbol{w}_i||} \text{の場合} \right) \\ 0 & \left(\text{それ以外の場合} \right) \end{cases}$$

パーセプトロン(5)

- 取れるマージンの大きさ

  ρ：クラス間マージン

$$\rho = \min_{x \in C_1} \frac{\boldsymbol{w}^T \boldsymbol{x}}{||\boldsymbol{w}||} - \max_{x \in C_2} \frac{\boldsymbol{w}^T \boldsymbol{x}}{||\boldsymbol{w}||}$$

最大マージンD_{max}

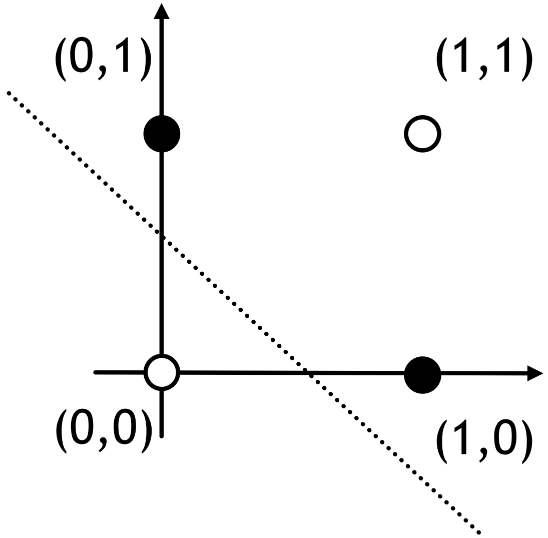
$$D_{max} = \frac{1}{2} \rho_{max}$$

誤差逆伝搬法(1)

- 多層パーセプトロン

例：排他的論理和関数

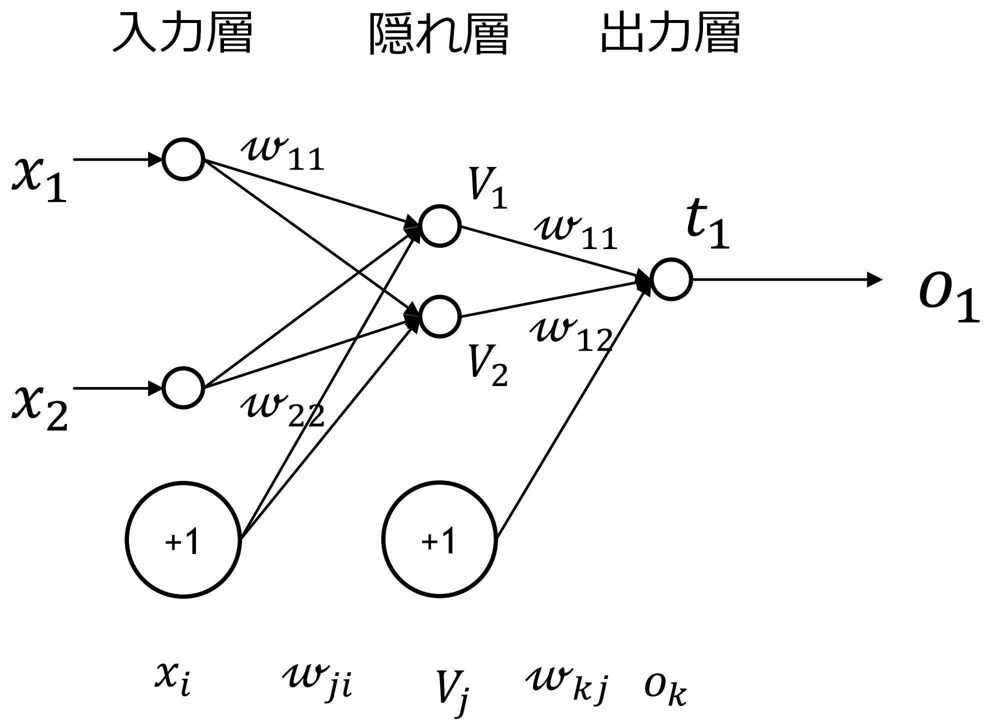
x_1	x_2	出力	教師データ
0	0	0	-1
0	1	1	+1
0	1	1	+1
1	1	0	-1



多層パーセプトロンの誤差逆伝搬法

誤差逆伝搬法(2)

- 多層パーセプトロンの誤差逆伝搬法



- V_jへの入力

$$h_j^n = \sum_{i=0}^{d} w_{ji} x_i^n = \boldsymbol{w}_j^T \boldsymbol{x}^n$$

- V_jからの出力

$$V_j^n = g(h_j^n)$$

誤差逆伝搬法(3)

- 非線形関数$g(u)$

  出力V_j^nが線形であれば、等価的に一層の回路で表現できてしまう

  - 多層回路である必要がない

$g(u)$を非線形とする

→**非線形出力関数**

例：シグモイド関数　$g(u) = \frac{1}{1+\exp(-\beta u)}$

誤差逆伝搬法(4)

- 出力o_k
  - $o_k (k = 1, \dots, K)$への入力

$$h_k^n = \sum_{j=0}^{M} w_{kj} V_j^n = \sum_{j=0}^{M} w_{kj} g\left(\sum_{i=0}^{d} w_{ji} x_i^n\right)$$

  - 出力

$$o_k^n = \tilde{g}(h_k^n) = \tilde{g}\left(\sum_{j=0}^{M} w_{kj} V_j^n\right) = \tilde{g}\left(\sum_{j=0}^{M} w_{kj} g\left(\sum_{i=0}^{d} w_{ji} x_i^n\right)\right)$$

誤差逆伝搬法 (5)

- 誤差逆伝搬法の学習規則

隠れ素子から出力素子への結合係数の学習

- 教師信号を用いて、二乗誤差最小化法を最急降下法に従って行う

$$E_n(\boldsymbol{w}) = \frac{1}{2}\sum_{k=1}^{K}(t_k^n - o_k^n)^2 = \frac{1}{2}\sum_{k=1}^{K}\left(t_k^n - \tilde{g}\left(\sum_{j=0}^{M} w_{kj} g\left(\sum_{i=0}^{d} w_{ji} x_i^n\right)\right)\right)^2$$

学習データ全体では、

$$E(\boldsymbol{w}) = \sum_{n=1}^{N} E_n(\boldsymbol{w})$$

誤差逆伝搬法(6)

- バッチアルゴリズム
  $E(\boldsymbol{w})$を評価関数として用いる
  学習データ全体を用いて結合係数の修正量を計算する
  →1エポック
  τエポック目の修正量を$\Delta w_{kj}(\tau)$とすると、、、

$$\Delta w_{kj}(\tau) = \sum_{n=1}^{N}\left(-\eta\frac{\partial E_n(\boldsymbol{w})}{\partial w_{kj}}\right) = -\eta\sum_{n=1}^{N}\left(\frac{\partial E_n(w)}{\partial o_k^n}\frac{\partial o_k^n}{\partial w_{kj}}\right)$$

誤差逆伝搬法(7)

- 修正量

$$\Delta w_{kj}(\tau) = \sum_{n=1}^{N}\left(-\eta\frac{\partial E_n(\boldsymbol{w})}{\partial w_{kj}}\right) = -\eta\sum_{n=1}^{N}\left(\frac{\partial E_n(w)}{\partial o_k^n}\frac{\partial o_k^n}{\partial w_{kj}}\right)$$

$$= \eta\sum_{n=1}^{N}(t_k^n - o_k^n)\,\tilde{g}(h_k^n)V_j^n = \eta\sum_{n=1}^{N}\delta_k^n V_j^n$$