パターン認識と学習

岡山大学大学院

講師　竹内孔一

本日の内容

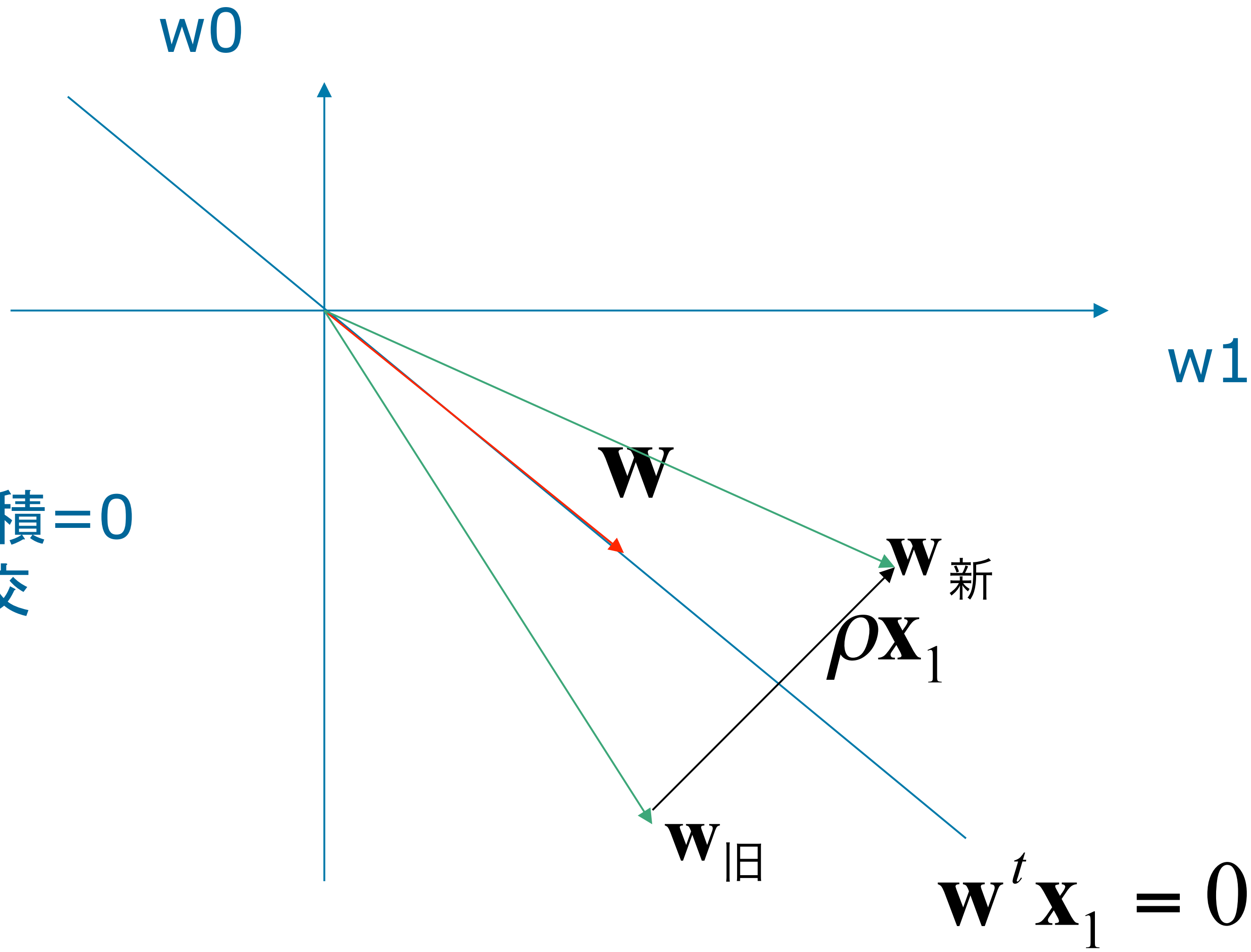
- 学習
  - Perceptron
  - Widrow-hoff learning-rule
    - 平均二乗誤差を最小

パーセプトロン

- 特徴
  - 簡易な手続きで表現
  - 線形分離可能ならば必ず解を見つける
  - 2値分類
  - 収束する理由(その1)
    - 誤ったときに正しい方向にρだけ進める

http://neuron.eng.wayne.edu/
学習モデルのいろいろ

直交について



wtx=0
なので内積=0
よって直交

学習規則

- 識別関数

$\boxed{g1(x)-g2(x)=w^t x}$

$g(x) = w^t x > 0 \quad x \in A1$

$g(x) = w^t x < 0 \quad x \in A2$

- 学習規則 (線形分離可能な場合のみ)

wの初期値を決める

  - 学習パターンを1つ選ぶ
  - $w' = w + \rho x$　(A1をA2と誤るとき)
  - $w' = w - \rho x$　(A2をA1と誤るとき)
  - 全パターンに対して繰り返す
  - 誤りがなくなるまで上記を繰り返す

例題

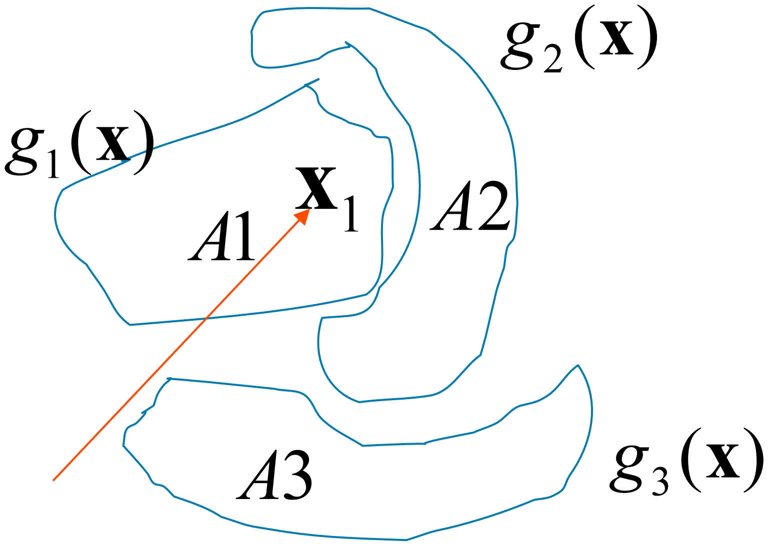
- w=(2,-6)は x1に対して誤るのでρ=1として修正した図を書け

Widrow-hoffの学習規則

- 何をするもの?
  - 識別関数を学習させる
- アイデア
  - 学習データxpに対して教師信号upを用意
  - 誤差を最小にする
- 得られるもの
  - 重回帰分析
  - widrow-hoff学習規則(パーセプトロンも含む)

教師信号と正解の違い

- 従来の正解(学習パターン) (x1, A1)
  - ベクトルと正解のセット
- 教師信号 $\mathbf{u}_p$
  - 各学習パターンについて識別関数でどう出力されたいか指定する(より強い指定)



これを指定

$$\begin{pmatrix} g_1(\mathbf{x}_1) = 1 \\ g_2(\mathbf{x}_1) = 0 \\ g_3(\mathbf{x}_1) = 0 \end{pmatrix} \rightarrow \mathbf{u}_1$$

カテゴリ数だけの要素がある(例は3カテゴリ)

練習4

- 次の誤差を求めなさい

識別関数 $g(\mathbf{x}_p)$ の値が $g(\mathbf{x}_p)=\begin{pmatrix}0\\1\\0\end{pmatrix}$

であるとき教師信号は $\mathbf{u}_p=\begin{pmatrix}1\\0\\0\end{pmatrix}$ であるとする

2つの解法

- 解析的手法
  - 誤差最小→ 誤差の式(2乗)を微分＝0
  - 直接解が求まる

  (欠点) 式は求まっても実際計算が無理
    - 大次元行列計算
- 数値解析による手法
  - 最急降下法
  - 逐次更新によって最小値を求める

練習5

- 式(A)が(B)に変換できることを確認せよ

$$\frac{1}{2}\sum_{p=1}^{N}\sum_{i=1}^{C}(\mathbf{w}_i^t\mathbf{x}_p - u_{ip})^2 \qquad \text{(A)}$$

$$\frac{1}{2}\sum_{i=1}^{C}\left\|\mathbf{X}\mathbf{w}_i - \mathbf{u}_{ip}\right\|^2 \qquad \text{(B)}$$

回答

まずwxの項に関して
入れ替えを考える

$$\sum_{p=1}^{n} \mathbf{w}_i^t \mathbf{x}_p = \begin{pmatrix} w_{i0} & \cdots & \cdots & w_{id} \end{pmatrix} \begin{pmatrix} x_0 \\ \vdots \\ \vdots \\ x_d \end{pmatrix}_{p=1} + \cdots\cdots + \begin{pmatrix} x_0 \\ \vdots \\ \vdots \\ x_d \end{pmatrix}_{p=n}$$

$$= \begin{pmatrix} x_0 & \cdots & \cdots & x_d \end{pmatrix}_{p=1} \begin{pmatrix} w_{i0} \\ \vdots \\ \vdots \\ w_{id} \end{pmatrix} + \ldots + \begin{pmatrix} x_0 & \cdots & \cdots & x_d \end{pmatrix}_{p=n} \begin{pmatrix} w_{i0} \\ \vdots \\ \vdots \\ w_{id} \end{pmatrix}$$ (k1)

(K1)は $\begin{pmatrix} \begin{pmatrix} x_0 & & x_d \end{pmatrix}_{p=1} \\ \\ \begin{pmatrix} x_0 & & x_d \end{pmatrix}_{p=n} \end{pmatrix} \begin{pmatrix} w_{i0} \\ \\ w_{id} \end{pmatrix} = \mathbf{X}\mathbf{w}_i$ の各要素を足し算したもので表せられる

(k2)

回答

次に全体で考えてみよう

$$\sum_{p=1}^{n}(\mathbf{w}_i^t\mathbf{x}_p - u_{ip})^2 = \underline{(\mathbf{w}_i^t\mathbf{x}_1 - u_{i1})^2 + \cdots + (\mathbf{w}_i^t\mathbf{x}_n - u_{in})^2}$$

この部分を行列で表そうとするとk2の関係を利用して　(k3)

$$\begin{pmatrix} (x_0 \quad\quad x_d)_{p=1} \\ \\ \\ (x_0 \quad\quad x_d)_{p=n} \end{pmatrix} \begin{pmatrix} w_{i0} \\ \\ \\ w_{id} \end{pmatrix} - \begin{pmatrix} u_{i\,p=1} \\ \\ \\ u_{i\,p=n} \end{pmatrix}$$

<=この各要素の
2乗の足し算したものが
上記の式(の意味)に等しい
なぜかは次のスライド参照

よって上式(k3)は

$$\sum_{p=1}^{n}(\mathbf{w}_i^t\mathbf{x}_p - u_{ip})^2 = \left\|\mathbf{X}\mathbf{w}_i - \mathbf{u}_i\right\|^2$$

となり，A式はB式に対応する

考察(誤差について)1/2

- 考察

$$\frac{1}{2}\sum_{p=1}^{N}\sum_{i=1}^{C}(\mathbf{w}_i^t\mathbf{x}_p - u_{ip})^2$$

$$e_{ip} = g_i(\mathbf{x}_p) - u_{ip}$$

この式の2乗のもともと意味は
ノルムの2乗という意味.
誤差の定義まで戻って
考えてみよう

実はこのeはとてもくせもの. e(誤差)はベクトルのノルム
というのが暗に定義されているのです. つまり $e = \|e\|$
なので先ほどの計算でも||Xw-b||という形にします

本当かどうか試してみましょう

識別関数があってる場合(a1)と間違ってる場合(a2)の
距離を実際にeで測って見ましょう

考察(誤差について)2/2

- 状況: 3つのクラス(A1,A2,A3)がありA2の学習データ x_1, x_2 の2つについてe(誤差)を測定する

$$e_{ip} = g_i(\mathbf{x}_p) - u_{ip}$$

g(x)　教師

正しい識別

$$e_1 = \begin{pmatrix} 0 & & 0 \\ 1 & - & 1 \\ 0 & & 0 \end{pmatrix}$$

誤識別

$$e_2 = \begin{pmatrix} 1 & & 0 \\ 0 & - & 1 \\ 0 & & 0 \end{pmatrix}$$

eをどう計算するか?

もし各要素を単純に引き算で計算すると

$$e_1 = 0 - 0 + 1 - 1 + 0 - 0 = 0$$
$$e_2 = 1 - 0 + 0 - 1 + 0 - 0 = 0$$

->とどっちも0となり誤りが測定できません

→ノルムで計算します

$$e_1 = \sqrt{0^2 + 0^2 + 0^2} = 0$$
$$e_2 = \sqrt{(1-0)^2 + (0-1)^2 + 0^2} = \sqrt{2}$$

誤差が測定可能!

行列に関する変換(補足)p.36

$$\sum_{p=1}^{n}(\mathbf{w}_i^t\mathbf{x}_p)\mathbf{x}_p = (w_{i0},..,w_{id})\begin{pmatrix} x_0 \\ \\ x_d \end{pmatrix}_{p=1}\begin{pmatrix} x_0 \\ \\ x_d \end{pmatrix}_{p=1} + ...$$

$$= (w_{i0}x_0 + .. + w_{id}x_d)_{p=1}\begin{pmatrix} x_0 \\ \\ x_d \end{pmatrix}_{p=1} + .. + (w_{i0}x_0 + .. + w_{id}x_d)_{p=n}\begin{pmatrix} x_0 \\ \\ x_d \end{pmatrix}_{p=n}$$

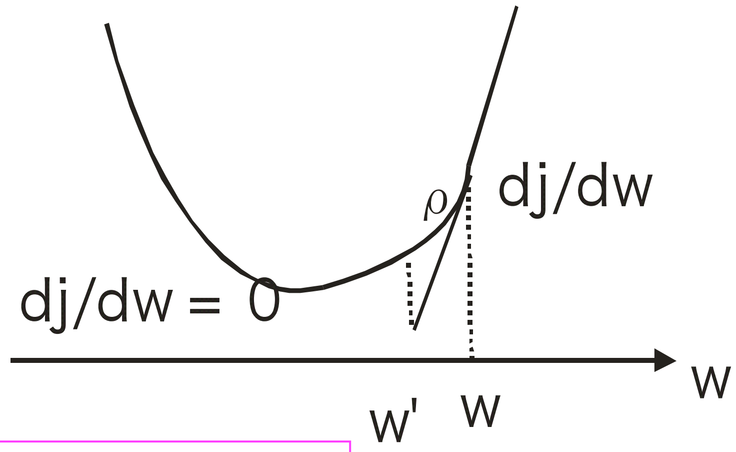
$$= \left(\begin{pmatrix} x_0 \\ \\ x_d \end{pmatrix}_{p=1} \cdots \begin{pmatrix} x_0 \\ \\ x_d \end{pmatrix}_{p=n}\right)\begin{pmatrix} (w_{i0}x_0 + .. + w_{id}x_d)_{p=1} \\ \vdots \\ (w_{i0}x_0 + .. + w_{id}x_d)_{p=n} \end{pmatrix}$$

$$= \left(\begin{pmatrix} x_0 \\ \\ x_d \end{pmatrix}_{p=1} \cdots \begin{pmatrix} x_0 \\ \\ x_d \end{pmatrix}_{p=n}\right)\begin{pmatrix} (x_0 & & x_d)_{p=1} \\ & \vdots & \\ (x_0 & & x_d)_{p=n} \end{pmatrix}\begin{pmatrix} w_{i0} \\ \\ w_{id} \end{pmatrix} = \mathbf{X}^t\mathbf{X}\mathbf{w}_i$$

最急降下法

- 逐次近似
  - 誤差最小値の獲得

$$\mathbf{w}' = \mathbf{w} - \rho \frac{\partial J}{\partial \mathbf{w}}$$



ある関数Jに対してwが変数
であるとき，wで偏微分した傾きに
ρだけ更新する　→ 凸関数なら極に収束