

第9回講義「パターン認識」

単層ニューラルネットワーク入門

[1] 確率密度関数 vs. 識別関数

[2] 線形識別関数

- 2クラス

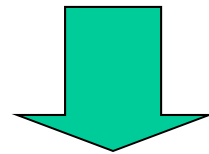
- 多クラス

- sigmoid関数と事後確率

誤り確率最小化による最適識別

$$P(C_k | \mathbf{x}) = \frac{p(\mathbf{x} | C_k) P(C_k)}{\sum_{j=1}^c p(\mathbf{x} | C_j) P(C_j)} \rightarrow \text{max for } C_k$$

事後確率



識別関数の構成

$$y_k(\mathbf{x}) = p(\mathbf{x} | C_k) P(C_k)$$

$y_j(\mathbf{x}) > y_k(\mathbf{x})$ for $\forall k \neq j \rightarrow \mathbf{x}$ is assigned to class C_j

$p(\mathbf{x} | C_k), P(C_k)$ を知らなくてはならない！

統計的パターン認識における 生成モデル vs. 識別モデル(1)

■ クラス分類問題における推論段階と決定段階

① 推論段階

学習データを用いて**事後確率** $P(C_k | \mathbf{x})$ を学習する。
そのため、**クラス条件付き確率密度** $p(\mathbf{x} | C_k)$ と
事前確率 $P(C_k)$ を推論する問題を扱う。

② 決定段階

観測ベクトル \mathbf{x} に対して、**事後確率** $P(C_k | \mathbf{x})$ が
最大となる最適なクラス C_j に割り当てを行う。

統計的パターン認識における 生成モデル vs. 識別モデル(2)

■ 統計的パターン認識における生成モデル

クラス条件付き確率密度 $p(x|C_k)$ と
事前確率 $P(C_k)$ のそれぞれを, 学習データを用いて, 直接, 推論する問題を解く。

■ 生成モデルの得失

1. x は高次元なので $p(x|C_k)$ を高精度で求めるには多くの学習データ(次元数の2桁以上)が必要となる。
2. 周辺分布 $p(x)$ から外れ値が検出できる。

統計的パターン認識における 生成モデル vs. 識別モデル(3)

■ 統計的パターン認識における識別モデル

推論と決定の段階を同時に扱い、観測ベクトル x を割り当てるクラスを直接決定する**識別関数**を学習する。理想的には**事後確率** $P(C_k | x)$ の直接推論になる。

■ 識別モデルの得失

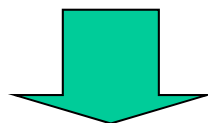
1. 学習データは比較的少なくても済む。ただし、決定領域の境界近傍データが必要となる。
2. リスク最小化や棄却オプションが扱えない。

識別関数の自由な導入

統計的パターン認識における識別モデルでは

- ・ 識別関数を直接構成する

ため, どのような関数形を選択するかが
学習および識別能力を左右する



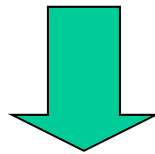
$$y_k = y_k(\mathbf{x}; \mathbf{w}), \quad \mathbf{w} : \text{parameters or weights}$$

\mathbf{w} は学習データから適当な学習アルゴリズムを用いて最適決定することになる

線形識別関数－2クラスの場合

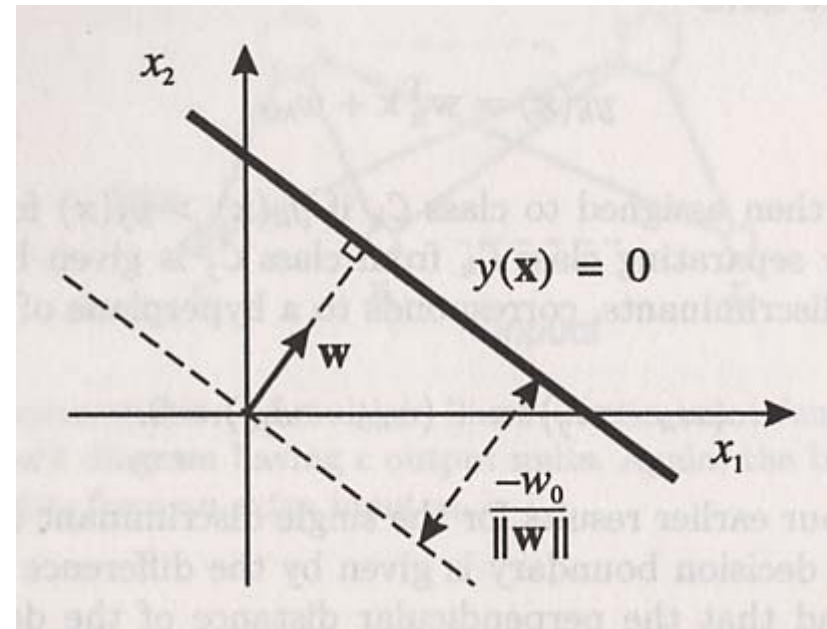
最も単純な識別関数形：

$$y(\mathbf{x}) = \mathbf{w}^t \mathbf{x} + w_0$$



\mathbf{x} is assigned to class C_1 if $y(\mathbf{x}) > 0$

\mathbf{x} is assigned to class C_2 if $y(\mathbf{x}) < 0$



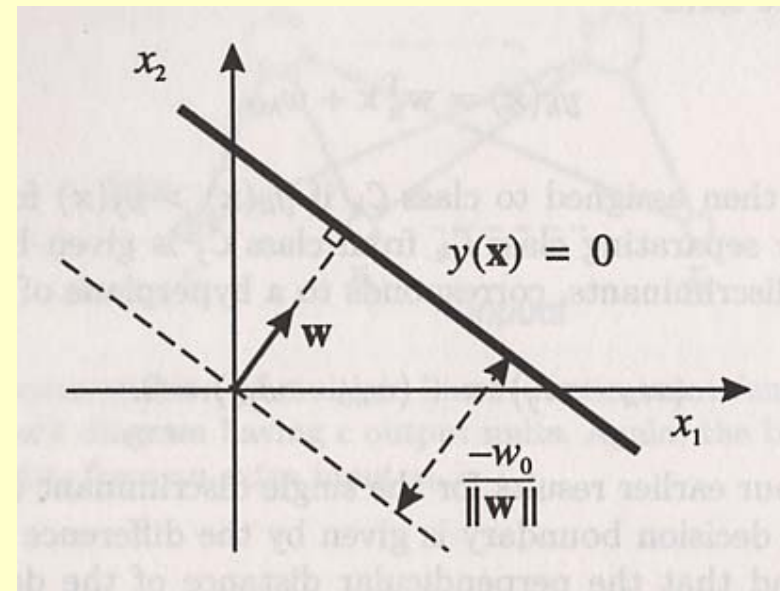
線形識別関数の決定境界

線形識別関数 $y(\mathbf{x}) = \mathbf{w}^t \mathbf{x} + w_0$ の決定境界は,

[1] ベクトル \mathbf{w} に垂直であり,

[2] 原点からの距離 l が
次式 で与えられる.

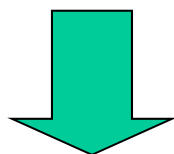
$$l = \frac{|w_0|}{\|\mathbf{w}\|}$$



線形識別関数と単層ニューラルネット

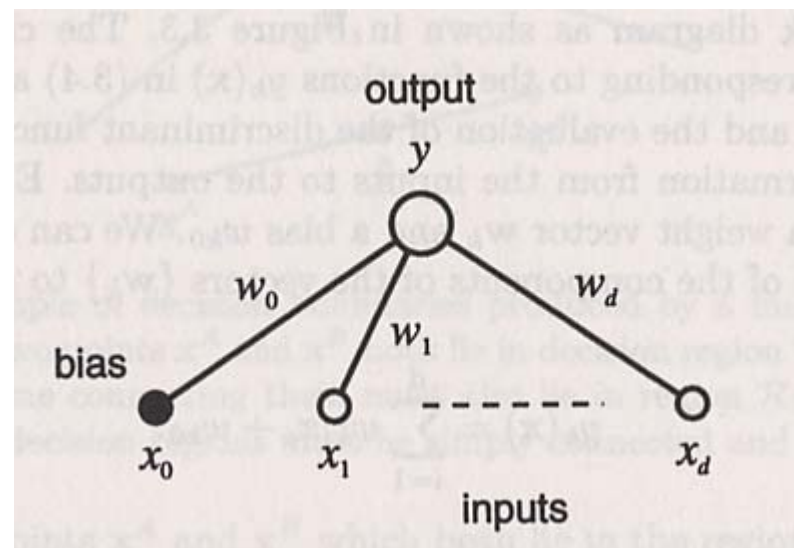
識別関数形の書き換え:

$$y(\mathbf{x}) = \mathbf{w}^t \mathbf{x} + w_0$$



$$y(\mathbf{x}) = \tilde{\mathbf{w}}^t \tilde{\mathbf{x}}$$

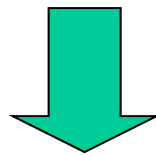
$$\tilde{\mathbf{w}} = (w_0, \mathbf{w}), \quad \tilde{\mathbf{x}} = (1, \mathbf{x})$$



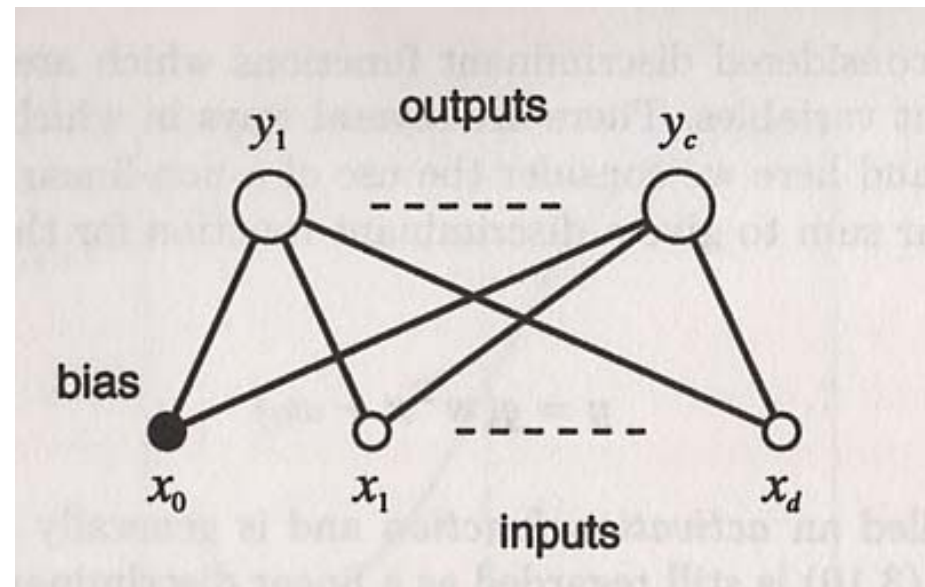
*weights*が単層を成している

線形識別関数－多クラスの場合

$$y_k(\mathbf{x}) = \mathbf{w}_k^t \mathbf{x} + w_{k0} = \sum_{i=1}^d w_{ki} x_i + w_{k0}$$
$$= \sum_{i=0}^d w_{ki} x_i \quad (1 \leq k \leq c), \quad x_0 = 1$$



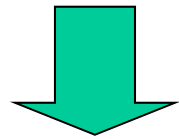
\mathbf{x} is assigned to class C_k
if $y_k(\mathbf{x}) > y_j(\mathbf{x})$
for $\forall j \neq k$



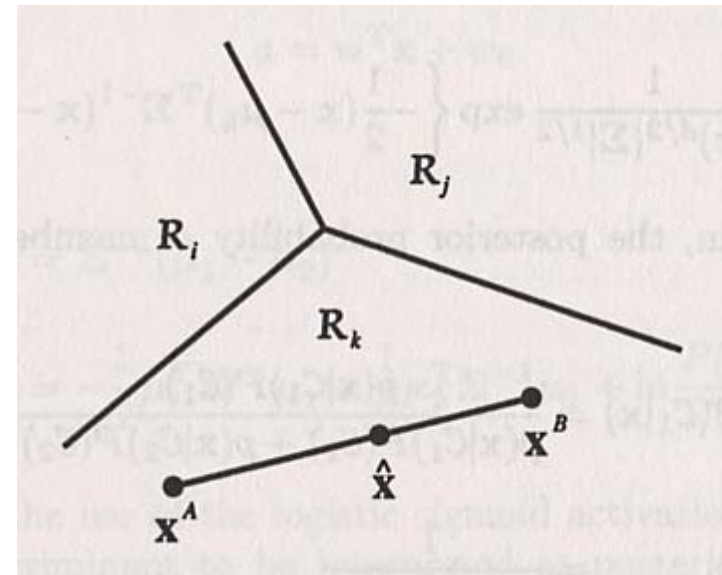
多クラス線形識別関数の決定領域

$$\hat{\mathbf{x}} = \alpha \mathbf{x}^A + (1 - \alpha) \mathbf{x}^B \quad (0 \leq \alpha \leq 1)$$

$$\text{if } \mathbf{x}^A \in \mathcal{R}_k \text{ and } \mathbf{x}^B \in \mathcal{R}_k \rightarrow \hat{\mathbf{x}} \in \mathcal{R}_k$$



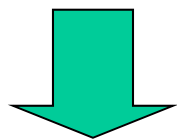
多クラスの決定領域
 $\{ R_1, \dots, R_k, \dots, R_c \}$ は
全て単連結で凸な領域
となる！



線形識別関数の発展形

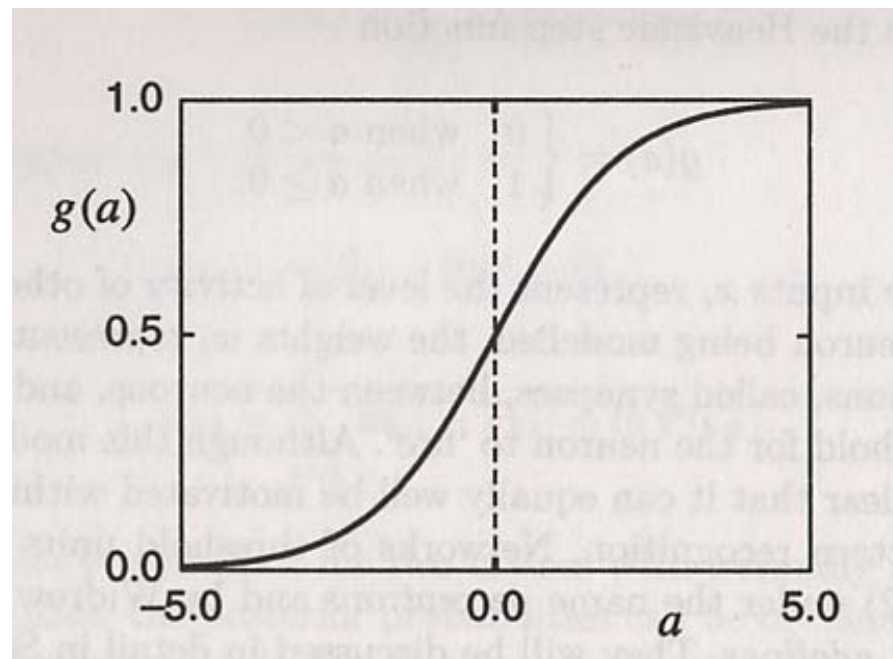
$$y = g(\mathbf{w}^t \mathbf{x} + w_0)$$

但し, $g(\bullet)$ は単調な非線形関数



特に, 次のsigmoid
関数が重要である.

$$g(a) = \frac{1}{1 + \exp(-a)}$$



sigmoid関数と事後確率の関係

2クラスの場合:

$$P(C_1 | \mathbf{x}) = \frac{p(\mathbf{x} | C_1) P(C_1)}{p(\mathbf{x} | C_1) P(C_1) + p(\mathbf{x} | C_2) P(C_2)}$$

$$= \frac{1}{1 + \exp(-a)} \quad \rightarrow \quad a = \ln \frac{p(\mathbf{x} | C_1) P(C_1)}{p(\mathbf{x} | C_2) P(C_2)}$$

特に, Σ が等しい2つの正規分布の場合を調べてみよう!

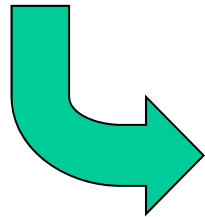
$$p(\mathbf{x} | C_k) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_k)^t \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_k) \right\}$$

sigmoid関数と事後確率の関係(続き)

$$a = \ln \frac{p(\mathbf{x} | C_1)P(C_1)}{p(\mathbf{x} | C_2)P(C_2)} = \mathbf{w}^t \mathbf{x} + w_0$$

$$\mathbf{w} = \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

$$w_0 = -\frac{1}{2} \boldsymbol{\mu}_1^t \Sigma^{-1} \boldsymbol{\mu}_1 + \frac{1}{2} \boldsymbol{\mu}_2^t \Sigma^{-1} \boldsymbol{\mu}_2 + \ln \frac{P(C_1)}{P(C_2)}$$



$$\therefore P(C_1 | \mathbf{x}) = g(\mathbf{w}^t \mathbf{x} + w_0)$$

➤ sigmoid関数は事後確率を表現できるから、識別関数として理想的となる

練習問題9-1

$$y_j(\mathbf{x}) = \mathbf{w}_j^t \mathbf{x} + w_{j0} \rightarrow \mathbf{x} \text{ is assigned to class } C_k$$
$$\text{if } y_k(\mathbf{x}) > y_j(\mathbf{x}) \text{ for } \forall j \neq k$$

により構成される決定領域 $\{ R_1, \dots, R_k, \dots, R_c \}$ は全て単連結で凸な領域となる. すなわち, 領域 R_k に属する任意の2点を結ぶ線分上の点も R_k に属する. これを証明するために次式が成り立つことを示せ.

$$\hat{\mathbf{x}} = \alpha \mathbf{x}^A + (1 - \alpha) \mathbf{x}^B \quad (0 \leq \alpha \leq 1)$$

$$\text{if } \mathbf{x}^A \in \mathfrak{R}_k \text{ and } \mathbf{x}^B \in \mathfrak{R}_k \rightarrow \hat{\mathbf{x}} \in \mathfrak{R}_k$$

練習問題9-2

$$p(\mathbf{x} | C_k) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_k)^t \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)\right\}$$

に従う2クラスの識別で, sigmoid関数を用いた線形識別関数と事後確率の間に次の関係が成り立つことを証明せよ.

$$P(C_1 | \mathbf{x}) = g(\mathbf{w}^t \mathbf{x} + w_0)$$

$$\mathbf{w} = \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

$$w_0 = -\frac{1}{2} \boldsymbol{\mu}_1^t \Sigma^{-1} \boldsymbol{\mu}_1 + \frac{1}{2} \boldsymbol{\mu}_2^t \Sigma^{-1} \boldsymbol{\mu}_2 + \ln \frac{P(C_1)}{P(C_2)}$$

練習問題9-2のヒント

$$-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^t \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_k)$$

などを展開するときには転置操作 t に関する性質:

$$(\mathbf{A} - \mathbf{B})^t = \mathbf{A}^t - \mathbf{B}^t, \quad (\mathbf{A}\mathbf{B})^t = \mathbf{B}^t \mathbf{A}^t, \quad (\mathbf{A}^t)^t = \mathbf{A}$$

および、数 a では $(a)^t = a$ となることを利用しなさい。

また、共分散行列の場合は $(\boldsymbol{\Sigma}^{-1})^t = \boldsymbol{\Sigma}^{-1}$ である。

従って、次のような式が成り立つ。

$$(\mathbf{x} - \boldsymbol{\mu}_k)^t = \mathbf{x}^t - \boldsymbol{\mu}_k^t$$

$$\boldsymbol{\mu}_k^t \boldsymbol{\Sigma}^{-1} = (\boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k)^t$$