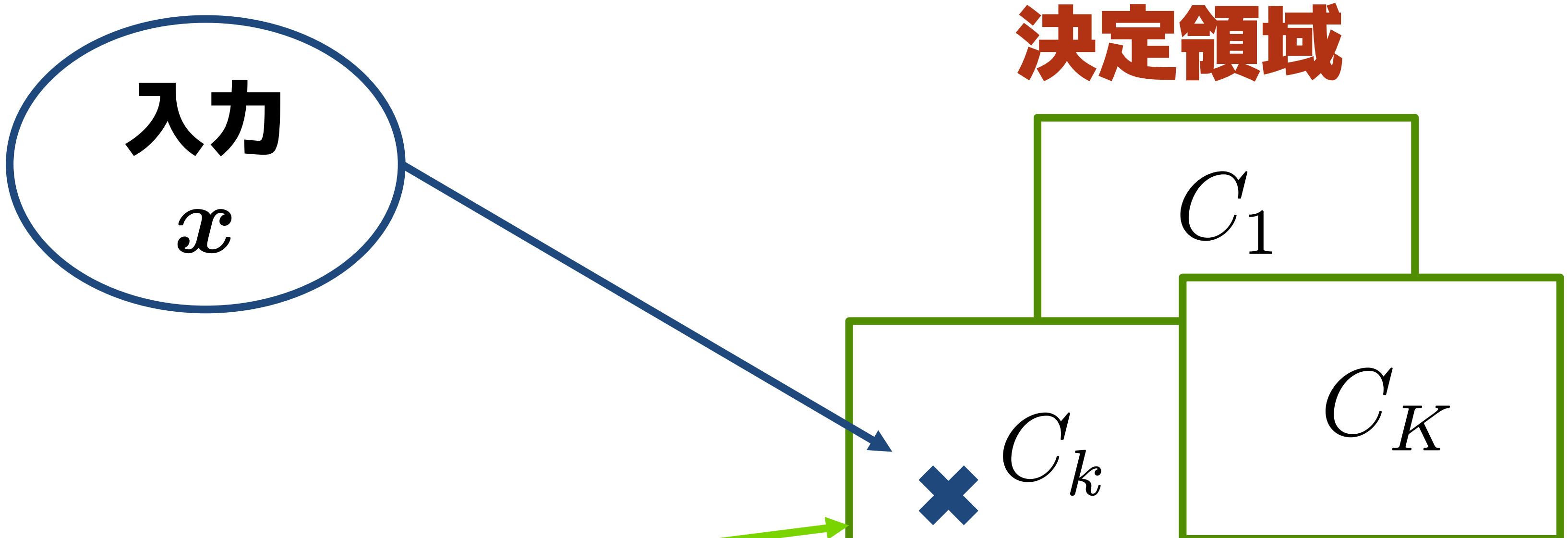
PRML輪講
第4章 線形識別モデル

東工大 理学院 物理学系
したろう

4章；分類問題

目的：入力をクラスの1つに割り当てる

決定領域



決定領域の境界

決定面 (or **決定境界**)

線形分離可能

線形な決定面で各クラスに分離できるデータ集合

分類問題のアプローチ；3択

- **識別関数を構築** **1.5.4 (c)**

  cf. パーセプトロン 4.1.7, SVM 7章

- **識別モデル** **1.5.4 (b)**

  事後確率$p(C_k|x)$を直接モデル化する方法

  cf. ロジスティック回帰 4.3.2

- **生成モデル** **1.5.4 (a)**

  1. 事前分布$p(x)$
     条件付き確率$p(x|C_k)$ の生成
  2. 事後確率$p(C_k|x)$を計算

$$p(C_k|\boldsymbol{x}) = \frac{\boxed{p(\boldsymbol{x}|C_k)p(C_k)}}{p(\boldsymbol{x})}$$

  cf. ナイーブベイズモデル 8.2.2

3章との比較

3章：線形回帰

入力 x $\xrightarrow[\text{予測}]{}$ 出力 t：**連続値**

$$y(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x} + w_0$$

線形な決定面

4章：一般化線形モデル

入力 x $\xrightarrow[\text{識別}]{}$ クラスラベル t：**離散値**

$$y(\boldsymbol{x}) = f(\boldsymbol{w}^T\boldsymbol{x} + w_0)$$

1 of K記法

クラスC_k

$\updownarrow$

$$t = (0, \ldots, 0, \overbrace{1}^{k\text{ 番目}}, 0, \ldots, 0)$$

非線形関数f：**活性化関数**

2クラスの識別 4.1.1

線形識別関数

バイアスパラメータ

$$y(\boldsymbol{x}) = \boxed{\boldsymbol{w}}^T \boldsymbol{x} + \boxed{w_0}$$

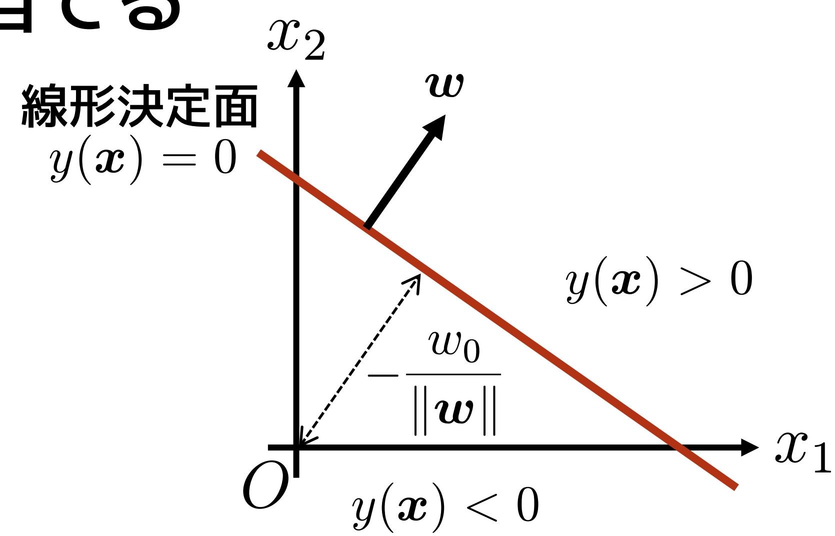
重みベクトル　　入力に対して**線形**

$y(\boldsymbol{x}) \geq 0$ →クラスC_1に割り当てる

otherwise →クラスC_2に割り当てる

ex. 2次元入力

$$\boldsymbol{w} = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$



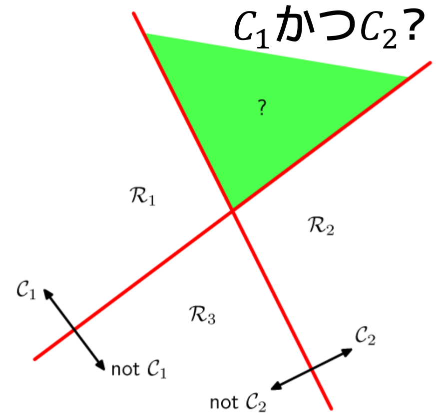
多クラスの識別へ拡張 4.1.2

方針：2クラス線形識別関数を組み合わせる

1. １対他分類器

$K-1$個の分類器を使う

クラスC_kである or NOT

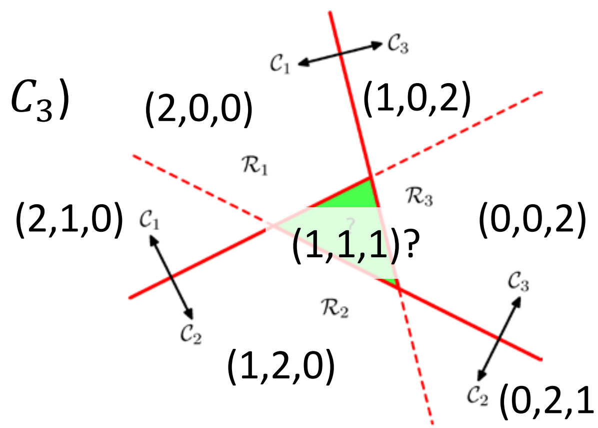


出典：http://research.microsoft.com/~cmbishop/PRML

2. １対１分類器

${}_K C_2$個のクラス対それぞれを2クラス分類器で分類

得票数:(C_1, C_2, C_3)



出典：http://research.microsoft.com/~cmbishop/PRML

いずれも
分類できない領域が生じる

多クラスの識別へ拡張 4.1.2

方針：単独のKクラス識別を考える

識別関数

$$y_k(\boldsymbol{x}) = \boldsymbol{w}_k^T \boldsymbol{x} + w_{k0}$$

識別方法

$$\bigwedge_{j(\neq k)} (y_k(\boldsymbol{x}) > y_j(\boldsymbol{x})) \Longrightarrow$$

入力xをクラスC_kに
割り当てる

cf. $y_1(x) > y_2(x) = 0$ なら
2クラスの識別に対応

決定面

$$(\boldsymbol{w}_k - \boldsymbol{w}_j)^T \boldsymbol{x} + (w_{k0} - w_{j0}) = 0$$

2クラスの決定面に対応

$$y(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} + w_0 = 0$$

決定領域

１つに連結していて**凸領域**

最小二乗法でパラメータを決定 4.1.2

入力　x：D次元
目標変数　t：1 of K記法

クラスC_kの線形モデル

ダミー入力$x_0 = 1$

$$y_k(\boldsymbol{x}) = \tilde{\boldsymbol{w}}_k^T \tilde{\boldsymbol{x}}_k \longleftarrow y(\boldsymbol{x})_k = \boldsymbol{w}_k^T \boldsymbol{x}_k + w_{k0}$$

入力xをクラスy_kが最大となるクラスに割り当てる

$$\tilde{\boldsymbol{x}}_k = \begin{pmatrix} x_0 \\ \boldsymbol{x}_k \end{pmatrix}, \ \tilde{\boldsymbol{w}}_k = \begin{pmatrix} w_{k0} \\ \boldsymbol{w}_k \end{pmatrix}$$

行列表示

$$\boldsymbol{y}(\boldsymbol{x}) = \tilde{W}^T \tilde{\boldsymbol{x}}$$

$$\tilde{W} = \begin{pmatrix} w_{10} & w_{20} & \cdots & w_{K0} \\ w_{11} & w_{21} & \cdots & w_{K1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1D} & w_{2D} & \cdots & w_{KD} \end{pmatrix}$$

Given：N個のデータ集合

最小二乗法

$$E_D(\tilde{W}) = \frac{1}{2} \| \tilde{X}\tilde{W} - T \|_F^2$$

$$T = \begin{pmatrix} \boldsymbol{t}_1^T \\ \vdots \\ \boldsymbol{t}_N^T \end{pmatrix}$$

$$\tilde{X}\tilde{W} = \begin{pmatrix} \tilde{\boldsymbol{x}}_1^T \tilde{\boldsymbol{w}}_1 & \cdots & \tilde{\boldsymbol{x}}_1^T \tilde{\boldsymbol{w}}_K \\ \vdots & \ddots & \vdots \\ \tilde{\boldsymbol{x}}_N^T \tilde{\boldsymbol{w}}_1 & \cdots & \tilde{\boldsymbol{x}}_N^T \tilde{\boldsymbol{w}}_K \end{pmatrix} = \begin{pmatrix} y_1(\boldsymbol{x})_1 & \cdots & y_K(\boldsymbol{x})_1 \\ \vdots & \ddots & \vdots \\ y_1(\boldsymbol{x})_N & \cdots & y_K(\boldsymbol{x})_N \end{pmatrix}$$

評価関数を最小化
→ $\mathrm{D}E_D(\tilde{W}) = 0$

$\| \cdot \|_F$：Frobeniusノルム

最小二乗法でパラメータを決定 4.1.2

評価関数

$$E_D(\tilde{W}) = \frac{1}{2}\|\tilde{X}\tilde{W} - T\|_F^2$$

$$DE_D(\tilde{W}) = 0$$

パラメータ

$$\tilde{W} = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T T = \tilde{X}^\dagger T$$

$$\boldsymbol{y}(\boldsymbol{x}) = \tilde{W}^T\tilde{\boldsymbol{x}}$$

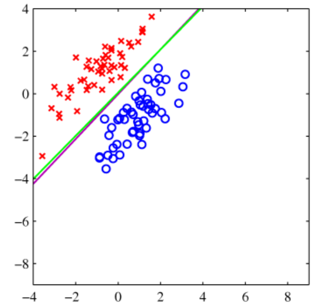
識別関数

$$\boldsymbol{y}(\boldsymbol{x}) = T^T(\tilde{X}^\dagger)^T\tilde{\boldsymbol{x}}$$

最小二乗法の問題点；2点 4.1.2

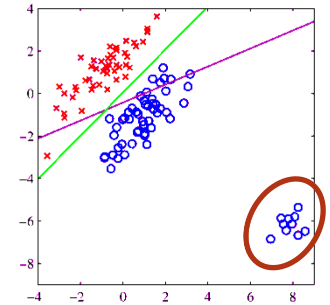
1. 外れ値に敏感

対策：別の誤差関数 7.1.2



出典：http://research.microsoft.com/~cmbishop/PRML

外れ値を追加



出典：http://research.microsoft.com/~cmbishop/PRML

2. 不適切な確率モデル

最小二乗法：Gauss分布を仮定

2値目的変数：Gauss分布に従っていない

$$t = (0, \ldots, 0, \overbrace{1}^{k\,\text{番目}}, 0, \ldots, 0)$$

対策：適切な確率モデルを採択する