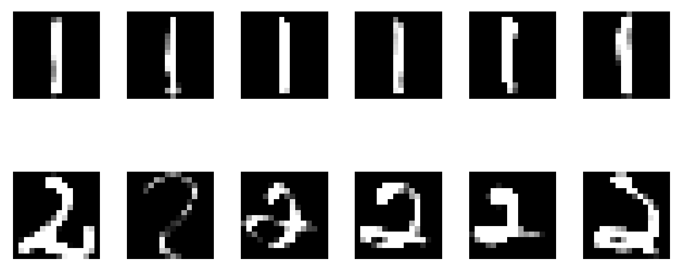
統計的機械学習
第2回　識別関数のよさを測る規準

2013/06/04

表現工学科　尾形研究室　ゼミ
野田　邦昭

手書き文字認識の例

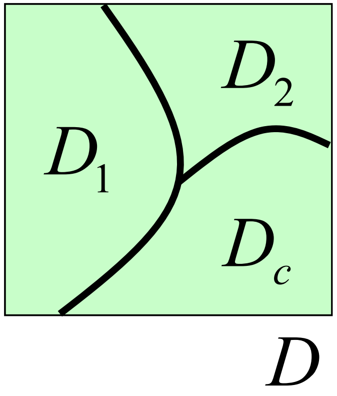
- スキャナで取り込んだ文字画像が16×16画素のとき，パターン x は各画素の濃度を縦に並べた256次元のベクトル.
- 厳密には画素値は実数ではない（例えば8ビット，即ち256階調の離散値）が，$[0,1]$ に正規化した実数値として扱う.
- このとき，パターン空間は $D=[0,1]^{256}$.
- カテゴリは各文字に対応.



識別関数・決定領域・決定境界

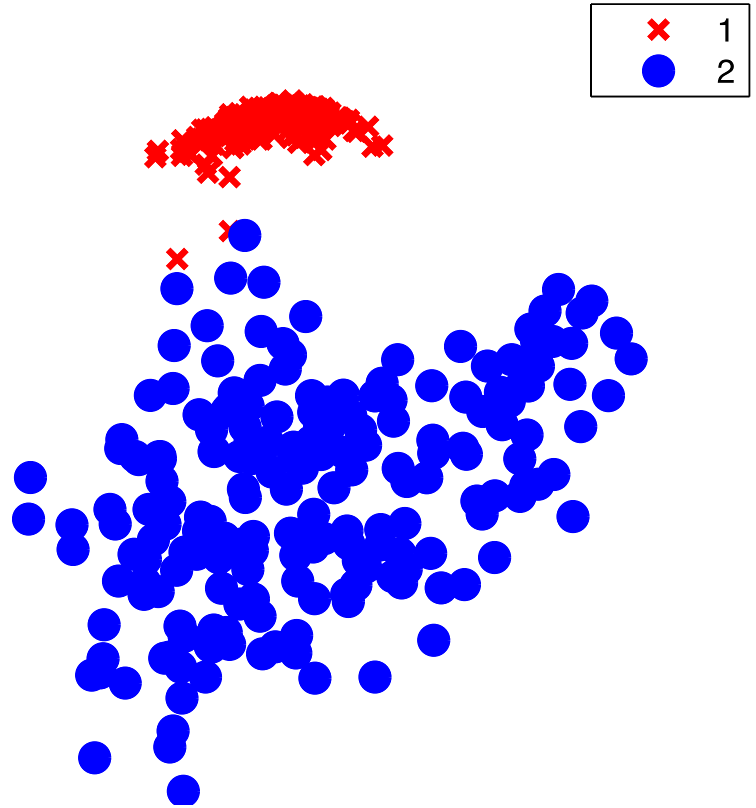
- 識別関数(discrimination function) $f(x)$: パターン xをそれが属するカテゴリ y に対応づける関数
- 決定領域(decision region) D_y : カテゴリ y のパターンが属する領域
- 決定境界(decision boundary) : いくつかの決定領域どうしの境界

識別関数を求めること
＝決定領域を求めること
＝決定境界を求めること

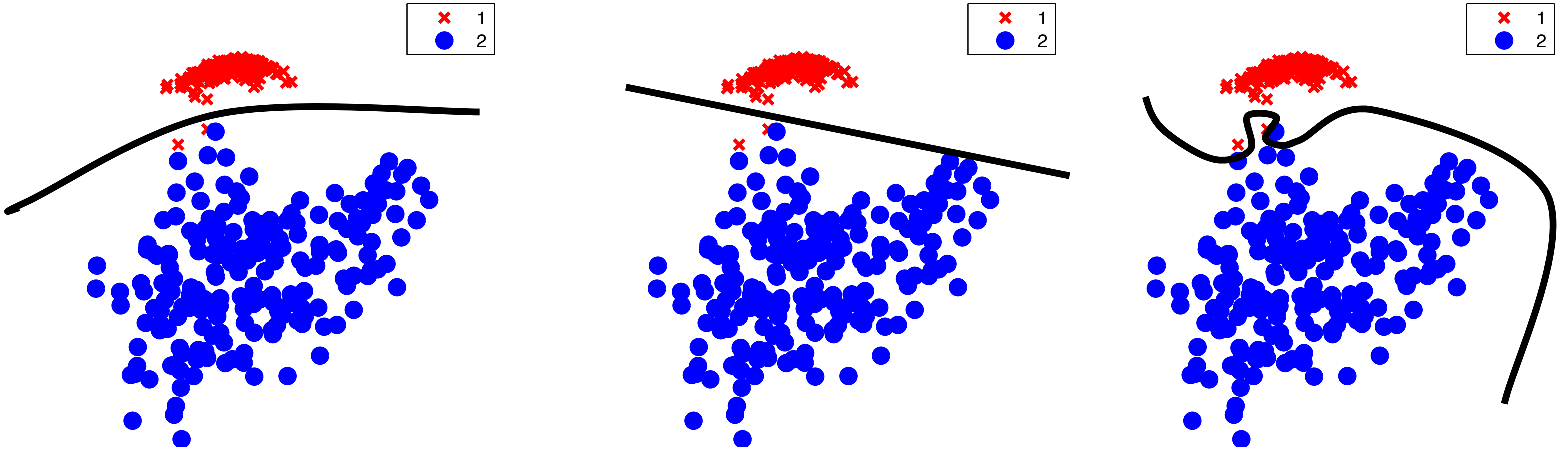


パターンの分布のイメージ

- 256次元空間内に分布しているパターンを適当な2次元部分空間に射影すると



どのような決定境界がよいか？



- 手持ちのパターンだけでなく，未知のパターンも正しく分類できるように，決定境界を定めたい．

識別関数のよさを測る規準

- よい識別関数を構成するためには，まず識別関数の「よさ」を測る規準が必要
  - 最大事後確率則
  - 最小誤識別率則
  - ベイズ決定則

最大事後確率則(1)

- 最大事後確率則(maximum a posteriori probability rule): 入力パターンが属する可能性が最も高いカテゴリを選ぶ

- これは, xを事後確率が最大になるカテゴリに分類することに対応:

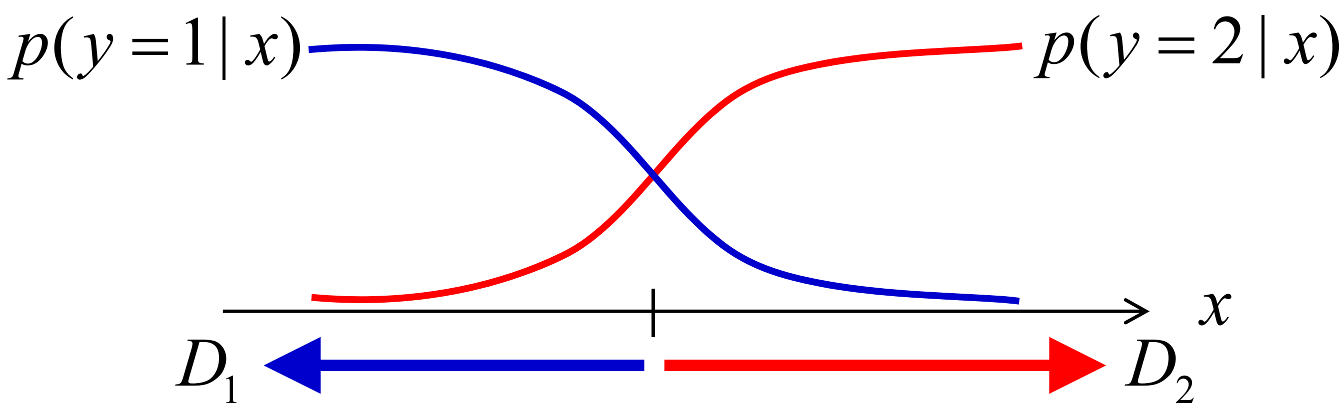
$$\arg\max_{y} \; p(y \mid x)$$

最大事後確率則(2)

$$\arg\max_{y} p(y \mid x)$$

■ 決定領域を次のように設定することとも等価:

$$D_y = \{x \mid p(y \mid x) \geq p(y' \mid x) \text{ for all } y' \neq y\}$$

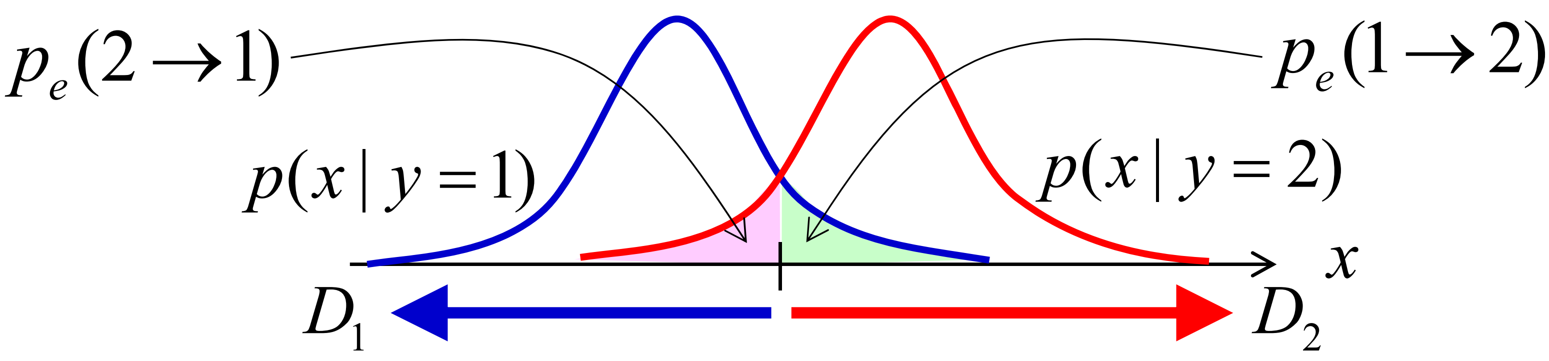


$p(y=1 \mid x) + p(y=2 \mid x) = 1$ （カテゴリ数 $c=2$ と仮定）

最小誤識別率則(1)

- 最小誤識別率則(minimum misclassification rate rule): パターンが誤って分類される確率を最小にするように識別関数を決定
- $p_e(y \to y')$:カテゴリ y に属するパターンが誤ってカテゴリ y' に分類される確率

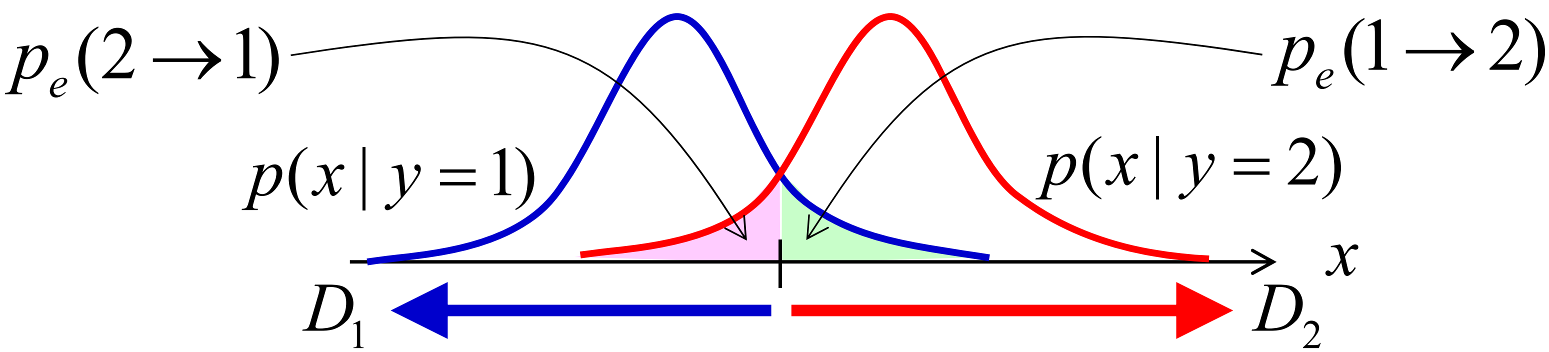
$$p_e(y \to y') = \int_{x \in D_{y'}} p(x \mid y)\,dx$$



最小誤識別率則（2）

$$p_e(y \to y') = \int_{x \in D_{y'}} p(x \mid y)dx$$

- これは，カテゴリ y に属するパターンが決定領域 $D_{y'}$ に入る確率と等価



最小誤識別率則(3)

- $p_e(y)$:カテゴリ y に属するパターンが誤って他のカテゴリに分類される確率

$$p_e(y) = \sum_{y' \neq y} p_e(y \to y')$$

- これは，以下のように分解できる:

$$\begin{aligned} p_e(y) &= \sum_{y' \neq y} \int_{x \in D_{y'}} p(x \mid y) dx \\ &\qquad + \int_{x \in D_y} p(x \mid y) dx - \int_{x \in D_y} p(x \mid y) dx \\ &= 1 - \underbrace{\int_{x \in D_y} p(x \mid y) dx}_{\text{正解率}} \end{aligned}$$

最小誤識別率則(4)

- 全体の誤識別率 p_e :

  $p_e(y)$ を全カテゴリーに対して平均したもの

$$p_e = \sum_{y=1}^{c} p_e(y) p(y)$$

- 最小誤識別率則では, p_e が最小になるように識別関数を決定する.
- 実は, 最小誤識別率則は最大事後確率則と等価である(証明は宿題).

誤識別と損失

- 最小誤識別率則に従えば，降水確率40%の時は雨が降らないと識別する．
- 雨が降らないならば傘を持っていく必要はないが，多くの人は降水確率40%ならば傘を持っていくであろう．
- それは，傘を持っていかなくて雨が降ったときの損失（雨にぬれて風邪をひく）が，傘を持っていって雨が降らなかったときの損失（かばんが少し重くなる）よりもずっと大きいからである．
- 宿題：他のおもしろい例を考えよ

ベイズ決定則(1)

- ベイズ決定則(Bayes decision rule): 誤って識別した時の損失を最小にするように識別
- $l_{y,y'}$:カテゴリ y に属するパターンを誤ってカテゴリ y' に分類したときの損失(loss)
- 条件付きリスク(conditional risk) $R(y' \mid x)$: パターン x をカテゴリ y' に分類したときの損失の期待値

$$R(y' \mid x) = \sum_{y=1}^{c} l_{y,y'} p(y \mid x)$$

ベイズ決定則(2)

$$R(y' \mid x) = \sum_{y=1}^{c} l_{y,y'} p(y \mid x)$$

- ベイズ決定則では，条件付きリスクが最小になるカテゴリにパターンを分類する

$$\underset{y}{\arg\min}\ R(y \mid x)$$

- これは，決定領域を次のように設定することと等価である.

$$D_y = \{x \mid R(y \mid x) \le R(y' \mid x) \text{ for all } y' \ne y\}$$

ベイズ決定則（３）

- 全リスク(total risk) R :条件付きリスクの全ての x に関する期待値

$$R = \int_D R(\hat{y} \mid x) p(x) dx$$

但し, $\hat{y}$ は識別機の出力を表す.

- ベイズリスク(Bayes risk):ベイズ決定則に対する全リスクの値

主成分分析 (PCA)