

情報認識

「識別関数のよさを測る規準」

- 担当教員： 杉山 将（計算工学専攻）
- 居室： W8E-505
- 電子メール： sugi@cs.titech.ac.jp

- 確率統計の復習
- 識別関数のよさを測る規準
- 条件付き確率の推定
 - パラメトリック法
 - 最尤推定法, EMアルゴリズム
 - ベイズ推定法, 最大事後確率推定法
 - ノンパラメトリック法
 - 核密度推定法
 - 最近傍密度推定法
- 手書き文字認識の計算機実習

パターンとカテゴリの表記

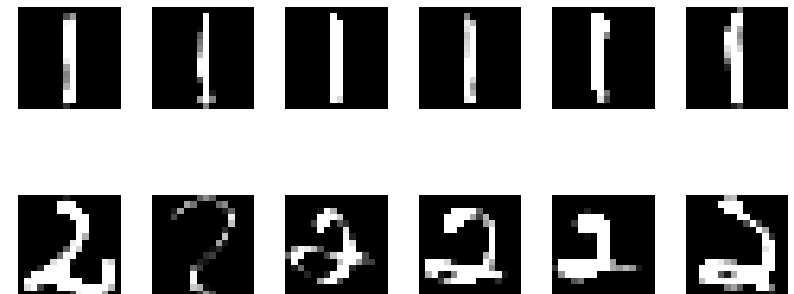
4

- パターン(pattern) x : d 次元実ベクトル
- パターン空間(pattern space) $D \left(\subset \mathbb{R}^d \right)$:
パターンの定義域(domain)
- y : カテゴリ(category) $y \in \{1, 2, \dots, m\}$
- m : カテゴリの数

手書き文字認識の例

5

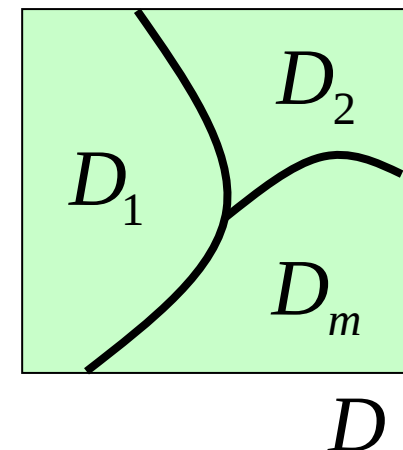
- スキャナで取り込んだ文字画像が 16×16 画素のとき, パターン x は各画素の濃度を縦に並べた256次元のベクトル.
- 厳密には画素値は実数ではない(例えば8ビット, 即ち256階調の離散値)が, $[0,1]$ に正規化した実数値として扱う.
- このとき, パターン空間は $D = [0,1]^{256}$.
- カテゴリは各文字に対応.



識別関数・決定領域・決定境界 6

- 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_i : 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

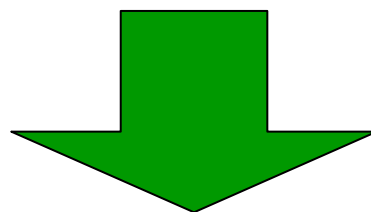
識別関数を求めること
= 決定領域を求めること
= 決定境界を求めること



統計的パターン認識

7

- 識別関数(決定領域, 決定境界)は未知



- 統計的パターン認識 (statistical pattern recognition): カテゴリ y やパターン x を確率変数として扱い, それらの統計的な性質を利用して識別関数を推定する

確率変数

8

- カテゴリ y やパターン x を確率変数(random variable)として扱えば, 次のような「確率」が定義できる.

$$p(x), p(y), p(x, y), p(y | x), p(x | y)$$

- カテゴリ y : 離散型(discrete type)の確率変数
- パターン x : 連続型(continuous type)の確率変数

確率関数と確率密度関数

9

- $p(y)$: カテゴリ y の生起確率を表す **確率関数**(probability function)

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

$$p(y) \geq 0 \text{ for } y = 1, 2, \dots, m$$

- $p(x)$: パターン x の **確率密度関数**(probability density function)

$$\int_D p(x) dx = 1$$

$$p(x) \geq 0 \text{ for all } x \in D$$

同時確率と条件付き確率

10

- $p(x, y)$: x と y の同時確率(joint probability)
- 周辺化(marginalization):

$$\sum_{y=1}^m p(x, y) = p(x)$$
$$\int_D p(x, y) dx = p(y)$$

周辺確率
(marginal probability)

- $p(x | y), p(y | x)$: 条件付き確率(conditional probability)

$$p(y | x)p(x) = p(x, y) = p(x | y)p(y)$$

事前確率・事後確率・ベイズの定理¹

- 事前確率(a priori probability) $p(y)$:
パターンを知る前のカテゴリの出現確率
- 事後確率(a posteriori probability) $p(y | x)$:
パターンを知った後のカテゴリの出現確率
- ベイズの定理(Bayes' theorem):

$$p(y | x) = \frac{p(x | y)p(y)}{p(x)}$$

期待値・分散共分散行列・独立性¹²

■ 期待値(expectation):

$$E[x] \equiv \int_D xp(x)dx$$

■ 分散共分散行列(variance covariance matrix):

$$V[x] \equiv E[(x - E[x])(x - E[x])^T]$$

■ x と x' が独立(independent):

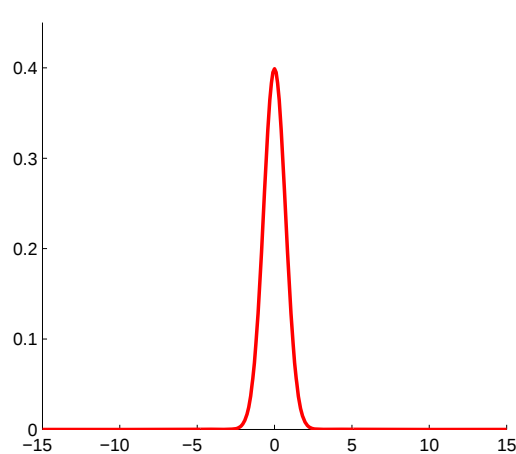
$$p(x, x') = p(x)p(x')$$

正規分布

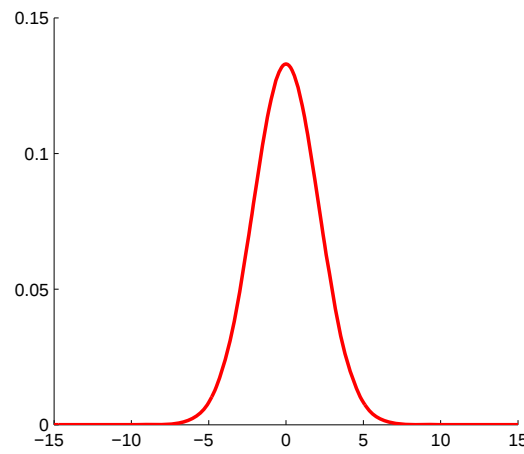
13

■ 2つのパラメータ: μ, σ^2

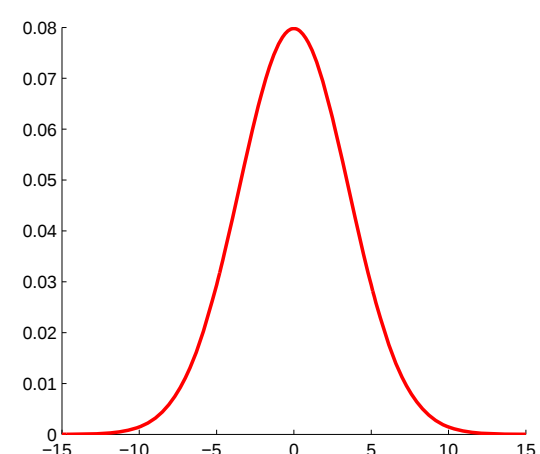
$$p(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$



$$\sigma^2 = 1$$



$$\sigma^2 = 9$$



$$\sigma^2 = 25$$

■ 正規分布の平均と分散:

$$E[x] = \mu$$

$$V[x] = \sigma^2$$

多次元正規分布

14

- d 次元の確率ベクトル: $x = (x^{(1)}, x^{(2)}, \dots, x^{(d)})^T$
- 2つのパラメータ:
 - d 次元ベクトル μ
 - d 次元正値行列 Σ

$$p(x; \mu, \Sigma) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$$

- 正規分布の期待値, 分散共分散行列

$$E[x] = \mu$$

$$V[x] = \Sigma$$

多次元正規分布(つづき)

15

- 共分散がゼロ (即ち $\Sigma = \text{diag}(\sigma_i^2)$) のとき

$\text{diag}(\sigma_i^2)$: $\sigma_1^2, \sigma_2^2, \dots, \sigma_d^2$ を
対角成分に持つ対角行列

$$p(x; \mu, \{\sigma_i^2\}_{i=1}^d) = \frac{1}{(2\pi)^{d/2} \prod_{i=1}^d \sigma_i} \exp\left(-\sum_{i=1}^d \frac{(x^{(i)} - \mu^{(i)})^2}{2\sigma_i^2}\right)$$

- さらに分散が等しい (即ち $\Sigma = \sigma^2 I$) のとき

I : 単位行列

$$p(x; \mu, \sigma) = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{(x - \mu)^T (x - \mu)}{2\sigma^2}\right)$$

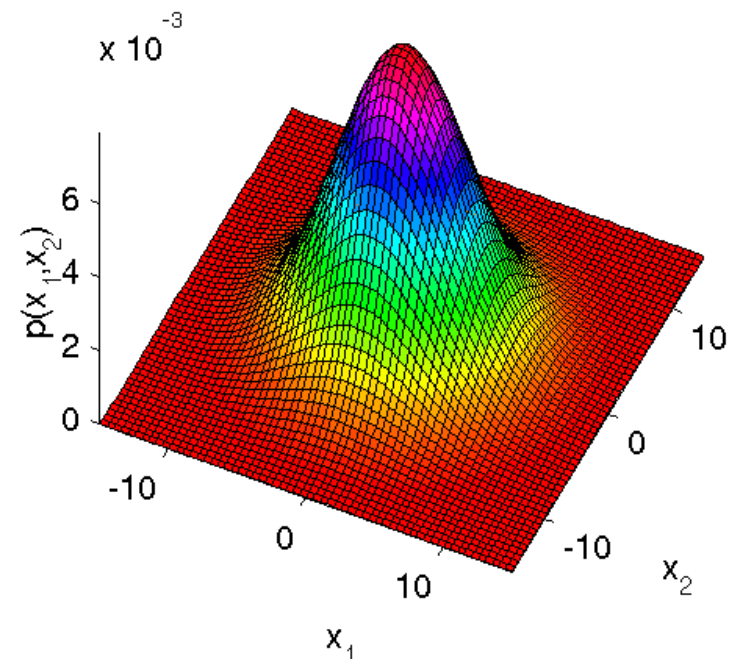
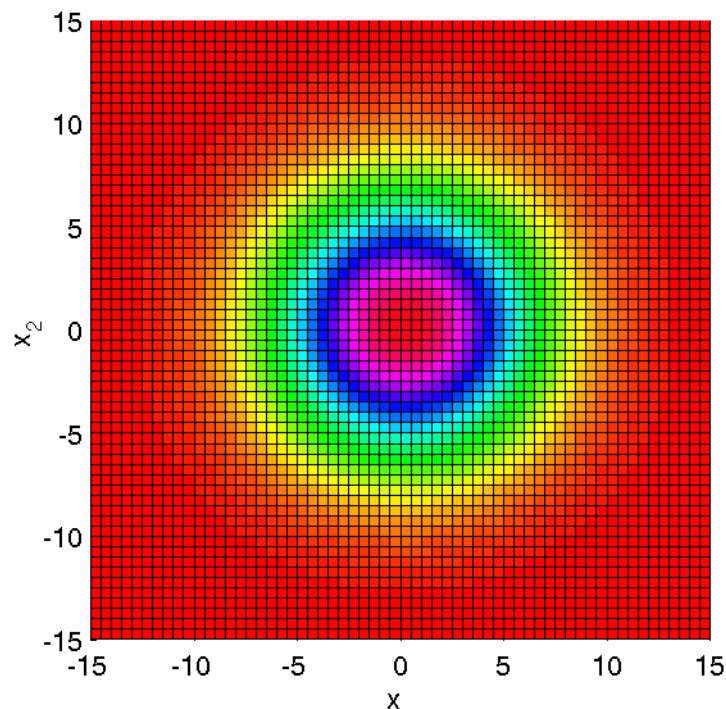
多次元正規分布の例(1)

16

$$d = 2$$

$$\mu = (0,0)^T$$

$$\Sigma = \begin{pmatrix} 20 & 0 \\ 0 & 20 \end{pmatrix}$$



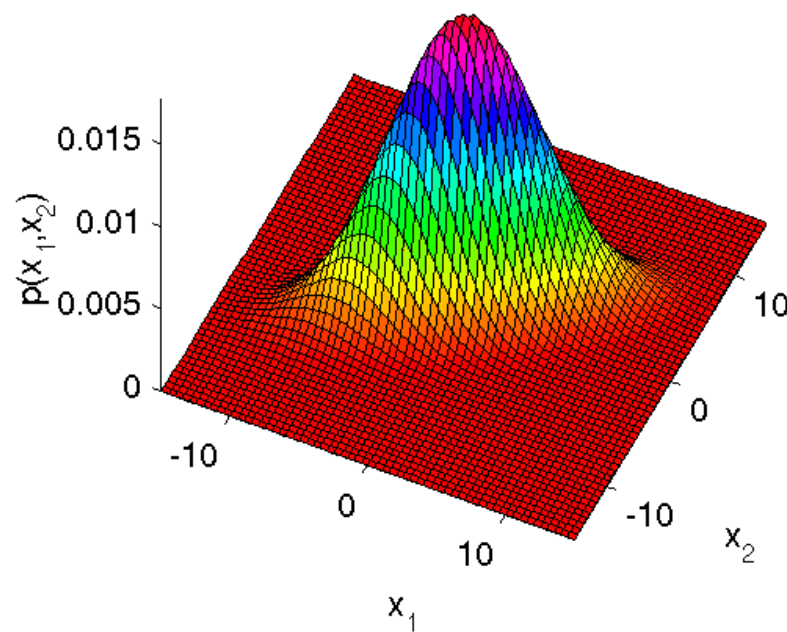
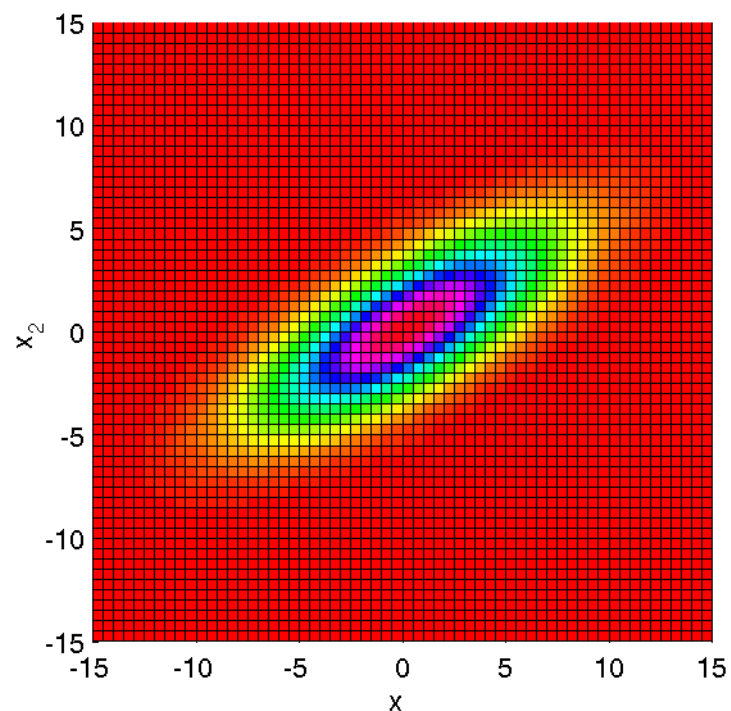
多次元正規分布の例(2)

17

$$d = 2$$

$$\mu = (0,0)^T$$

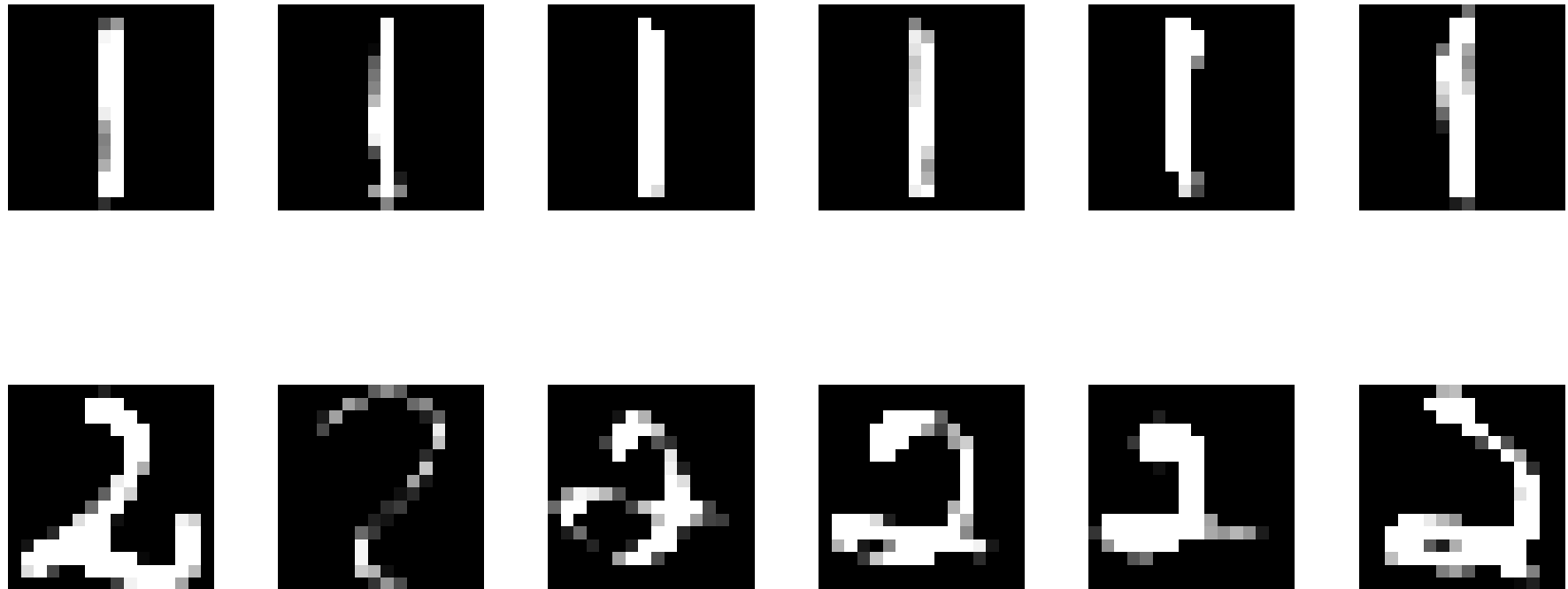
$$\Sigma = \begin{pmatrix} 20 & 10 \\ 10 & 9 \end{pmatrix}$$



手書き文字の例

18

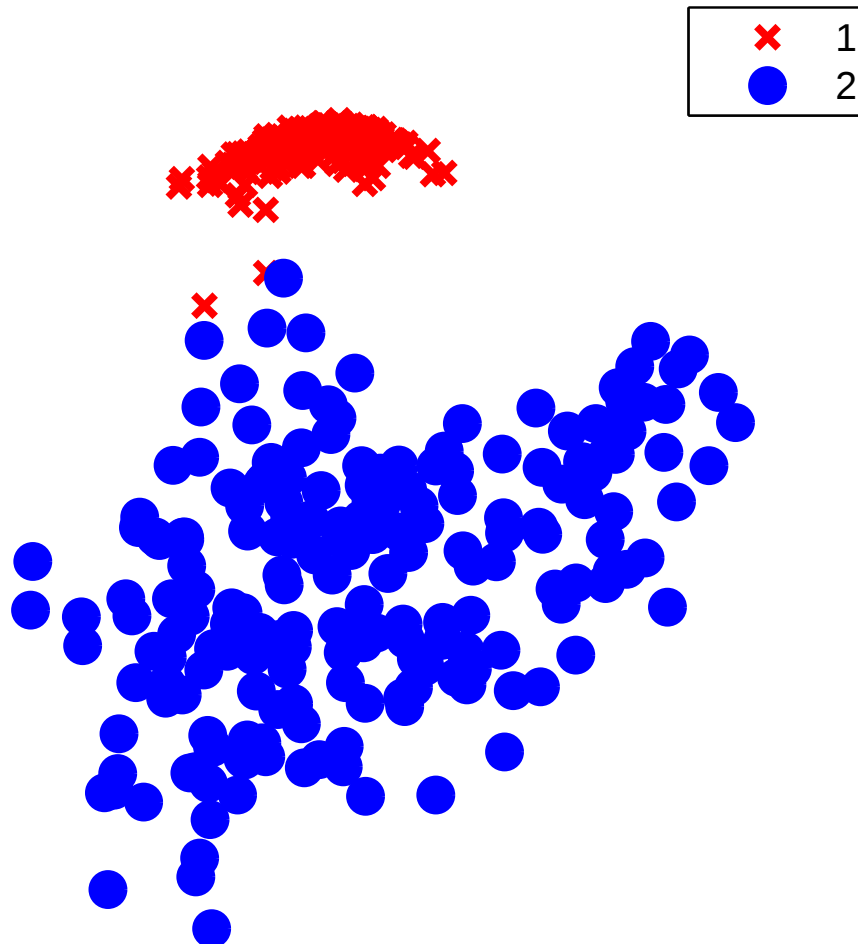
- 16 × 16画素, 各画素の濃度は0から255



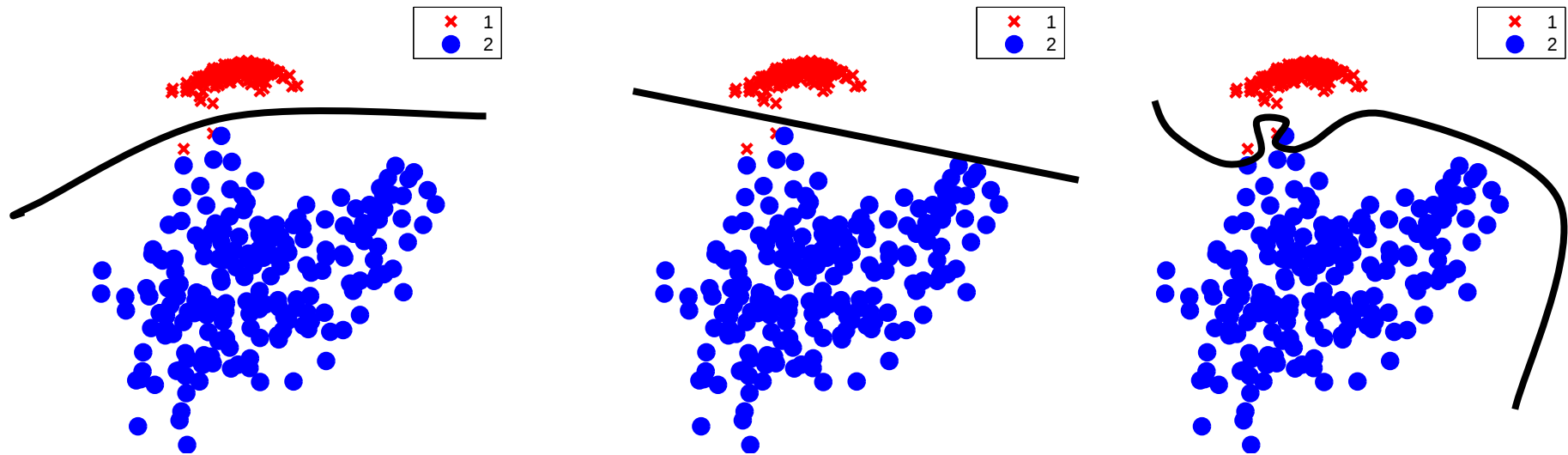
パターンの分布のイメージ

19

- 256次元空間内に分布しているパターンを適当な2次元に射影すると



どのような決定境界がよい？ 20



- 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

21

- 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

22

- 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する.

$$\arg \max_y p(y | x)$$

- また, 決定領域を次のように設定することとも等価である.

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

最小誤識別率則(1)

23

- 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
- $p_e(y \rightarrow y')$: 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

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

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

最小誤識別率則(2)

24

- $p_e(y)$: 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

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

- これは, 以下のように分解できる:

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

最小誤識別率則(3)

25

- 全体の誤識別率 p_e :

$p_e(y)$ を全カテゴリに対して平均したもの

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

- 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).

- 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、傘を持って行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)

27

- **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)**
- **条件付き危険(conditional risk)** $R(y' | x)$: 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

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

ベイズ決定則(2)

28

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

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

$$\arg \min_y R(y | x)$$

- 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.

$$D_y = \{x \mid R(y | x) \leq R(y' | x) \text{ for all } y' \neq y\}$$

ベイズ決定則(3)

29

- 全危険(total risk) R : 条件付き危険の全ての x に関する期待値

$$R = \int_D R(y' | x) p(x) dx$$

但し, y' は識別機の出力を表す.

- ベイズ危険(Bayes risk): ベイズ決定則に対する全危険の値



まとめ

30

- 識別関数のよさを測る3つの規準：
最大事後確率則，最小誤識別率則，ベイズ決定則
- 最大事後確率則と最小誤識別率則は等価
(証明は宿題！)。
- 損失が一定のベイズ決定則は最大事後確率則(及び最小誤識別率則)と等価(自明なので各自で確認せよ)。
- ベイズ決定則を用いるのが自然だが，現実には損失の値がはっきりしなかったり，計算が複雑になるといった理由から，最大事後確率則を用いることも多い。

小レポート(第2回)

31

1. Octaveなどを使い, 2次元正規分布の確率密度関数の3次元プロットおよび等高線プロットを, **3種類の異なる分散共分散行列**に対して作成せよ.
2. 分散共分散行列を変化させると分布の形がどのように変化するか論ぜよ.

ヒント: 共分散行列の固有方程式

$$\Sigma\phi = \lambda\phi$$

を解き Σ を固有値分解せよ.

$$\Sigma = \lambda_1\phi_1\phi_1^\top + \lambda_2\phi_2\phi_2^\top$$

- Octaveの使い方は, 添付の資料を参照せよ.

3. 最小誤識別率則によって得られる識別規則は、最大事後確率則によって得られるものと一致することを示せ
4. 誤って識別した場合の損失がカテゴリによって異なるようなパターン認識の実例を考えよ。また、それらの例では、損失の値は具体的にいくら位になるであろうか？

Octaveのサンプルプログラム

33

ex2.m

```
clear all
Mu=[0;0]; Sigma=[2 1;1 2];
x=[-3:0.1:3]; y=[-3:0.1:3];
for xx=1:length(x)
  for yy=1:length(y)
    z(xx,yy)=g2_pdf(x(xx),y(yy),Mu,Sigma);
  end
end

figure(1); clf
surf(x,y,z); view(45,60)
print -deps gauss2d_pdf_surf.eps

figure(2); clf
contour(x,y,z);
print -deps gauss2d_pdf_contour.eps
```

g2_pdf.m

```
function z=g2_pdf(x,y,Mu,Sigma)

d=sqrt(det(Sigma));
v=[x;y]-Mu;
z=1/(2*pi*d)*exp(-1/2*v'*inv(Sigma)*v);
```

実行例

34

