パターン認識

http://www.cs.miyazaki-u.ac.jp/~date/lectures/pattern/

伊達 章

宮崎大学 工学部 情報システム工学科

2017 年 10 月 25 日

講義のスケジュール（案）

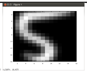
1. 講義の概要 10/5
2. 準備：確率・統計の基礎 10/12
3. 準備：octave の使い方 10/19
4. 教師あり学習．識別関数 10/26
5. 最大事後確率則，最小誤識別則，ベイズ決定則 11/2
6. 最尤推定法 1：ガウスモデル 11/9
7. 最尤推定法 2：線形判別分析 11/16
8. 線形判別分析により手書き文字認識 1
9. 線形判別分析により手書き文字認識 2
10. 混合ガウスモデルの最尤推定 1
11. 混合ガウスモデルの最尤推定 2
12. ノンパラメトリックな手法 (1)：カーネル密度推定法
13. ノンパラメトリックな手法 (2)：k-最近傍則
14. ノンパラメトリックな手法 (3)：パーセプトロン
15. 定期試験，解説

教師あり学習

パターン認識の問題

識別関数 $f(\boldsymbol{x})$ を作ること

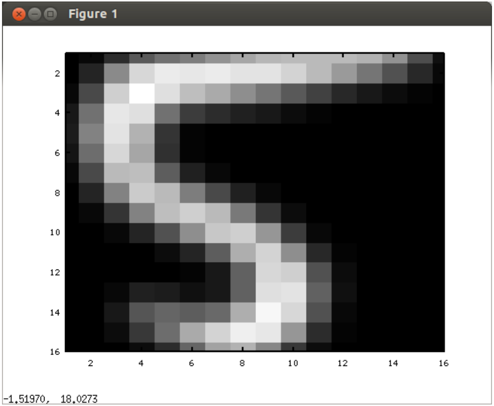
$$\boldsymbol{x} = (x_1, x_2, \cdots, x_{256}) \to y = f(\boldsymbol{x})$$

	$\boldsymbol{x}$	y
0	00$\cdots$00000000	$f(\boldsymbol{x}_0)$
1	00$\cdots$00000001	$f(\boldsymbol{x}_1)$
2	00$\cdots$00000010	$f(\boldsymbol{x}_2)$
3	00$\cdots$00000011	$f(\boldsymbol{x}_3)$
	$\vdots$	
k	00$\cdots$11101011 	$f(\boldsymbol{x}_k) = 5$
	$\vdots$	
$100\cdots0$	11 $\cdots$ 1111111	$f(\boldsymbol{x}_{10000\cdots0})$

パターン認識

教室付き学習

'5' → 生成



→ y: '5'
認識

訓練標本（例題）： $(\boldsymbol{x}_1, y_2), (\boldsymbol{x}_1, y_2), \cdots, (\boldsymbol{x}_n, y_n)$

主成分分析

主成分分析

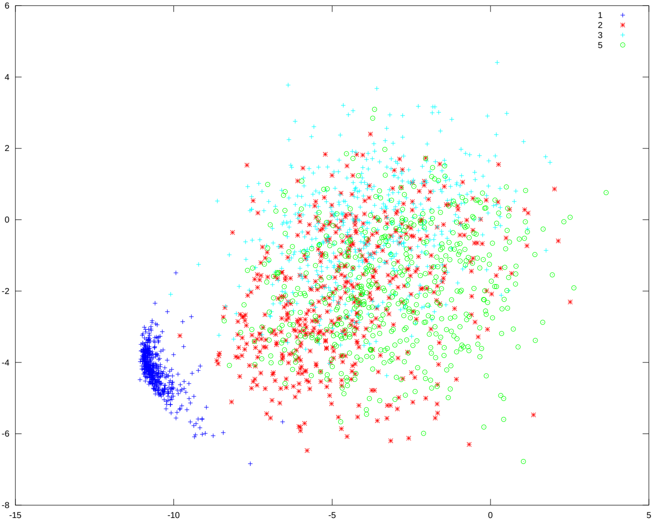
$$
\begin{aligned}
V &= \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^{\mathrm{T}},\ \boldsymbol{\mu}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i \\
V\boldsymbol{e}_j &= \lambda_j\boldsymbol{e}_j,\ \|\boldsymbol{e}_j\|=1,\ j=1,2,\cdots,256
\end{aligned}
$$

- V はデータの分散共分散行列．対称，正定値
- $n\cdots$ 例題の数.
- $\lambda_1>\lambda_2>\lambda_3>\cdots>0$
- $\boldsymbol{e}_1\cdots$ 第 1 主成分，$\boldsymbol{e}_2\cdots$ 第 2 主成分.
- $(\boldsymbol{e}_1\cdot\boldsymbol{x}_i,\boldsymbol{e}_2\cdot\boldsymbol{x}_i),\ i=1,\cdots,n$ （図 3.2）

$$
\boldsymbol{x} \approx \sum_{i=1}^{10}(\boldsymbol{e}_i\cdot\boldsymbol{x})\boldsymbol{e}_i
$$

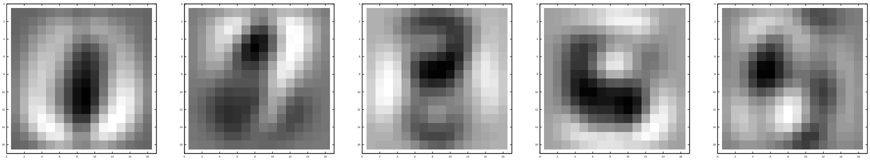
手書き文字データの分布（主成分分析）

$(\boldsymbol{e}_1 \cdot \boldsymbol{x}_i, \boldsymbol{e}_2 \cdot \boldsymbol{x}_i),\ i = 1, \cdots, 500(\times 4)$ （図 3.2）

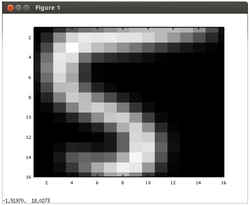


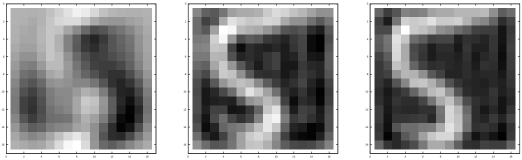
手書き文字データの主成分

$$\boldsymbol{e}_1 \sim \boldsymbol{e}_5$$



$$\boldsymbol{x} \approx \sum_{i=1}^{10,50,100} (\boldsymbol{e}_i \cdot \boldsymbol{x})\boldsymbol{e}_i$$





識別関数の良さを測る基準

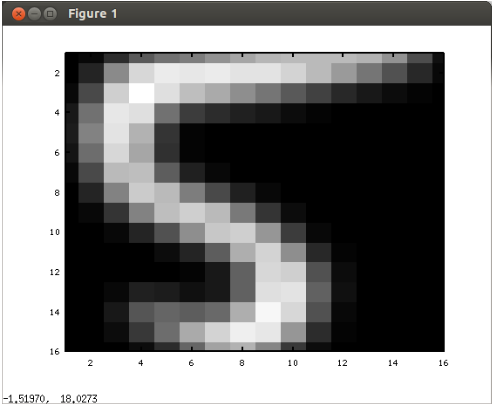
- 訓練標本を用いた識別関数の学習
- 最大事後確率則：$\hat{y} = \underset{y}{\operatorname{argmax}}\, p(y|\boldsymbol{x})$
- 最小誤識別率則
- ベイズ決定則： 条件付きリスク最小則

$$\hat{y} = \underset{y}{\operatorname{argmin}}\, R(y|\boldsymbol{x})$$

- 生成モデルに基づくパターン認識

確率的生成モデルに基づくパターン認識

'5' →
生成



→ y: '5'
認識

$$y \sim p(y) \rightarrow \text{データ}\ \boldsymbol{x} \sim p(\boldsymbol{x}|y) \rightarrow \text{認識}\ \hat{y} = \operatorname*{argmax}_{y} p(y|\boldsymbol{x})$$

モデル化： $p(y)$ と $p(\boldsymbol{x}|y)$ を設計する！

終