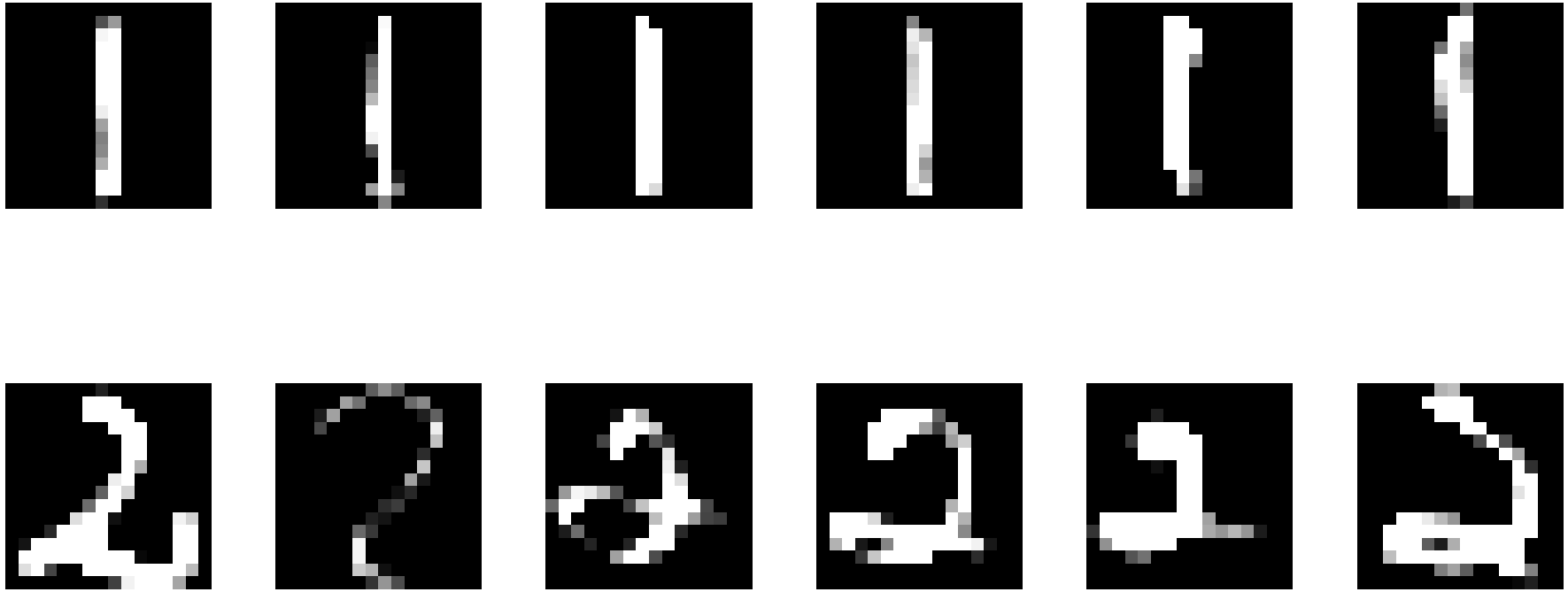
情報認識
「識別関数のよさを測る規準」

- 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

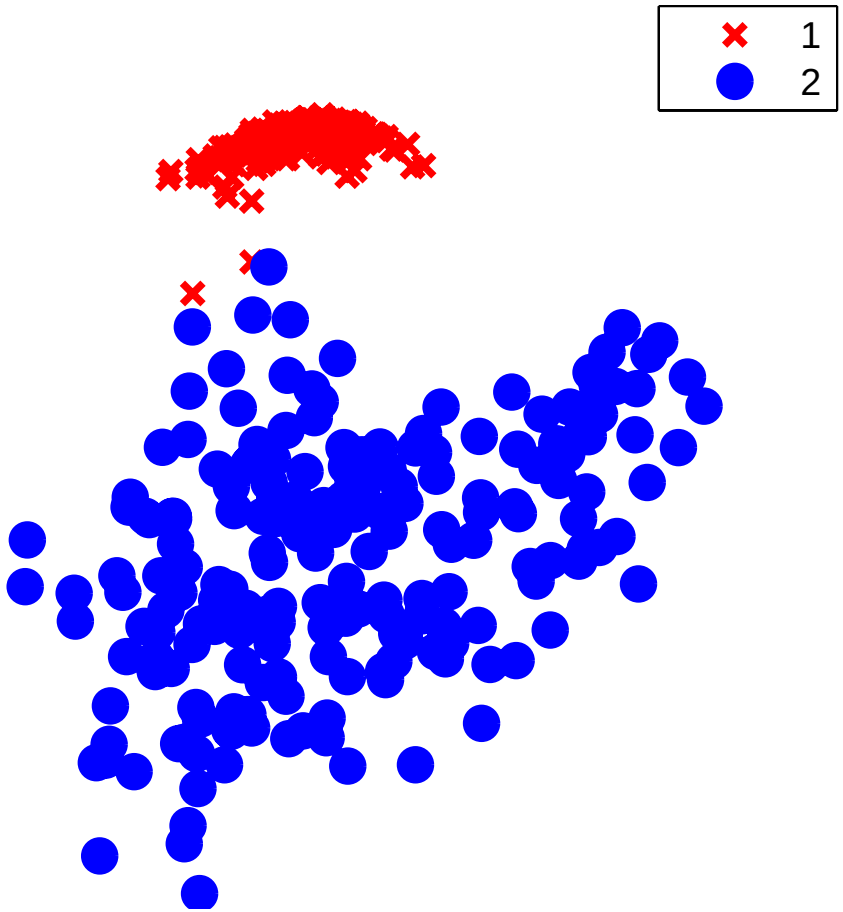
手書き文字の例

- １６×１６画素，各画素の濃度は０から２５５

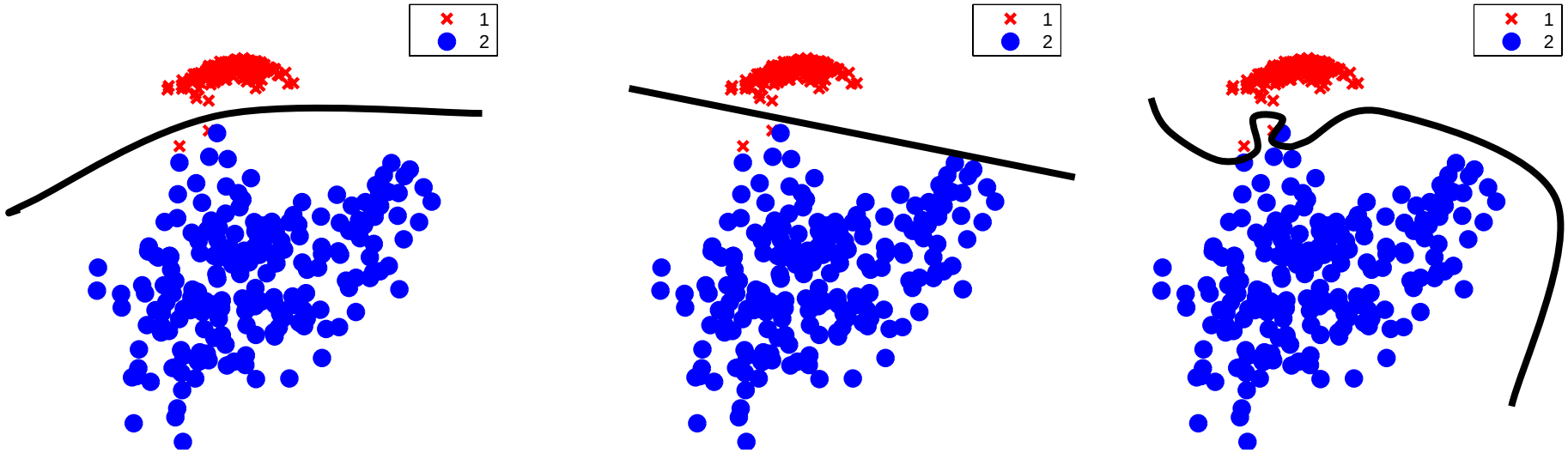


パターンの分布のイメージ

- 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，事後確率が最大になるカテゴリに分類することに対応.**

$$x \in \omega_k : k = \arg\max_i p(\omega_i \mid x)$$

- **決定領域を次のように設定することと等価.**

$$D_i = \{x \mid p(\omega_i \mid x) \geq p(\omega_j \mid x) \text{ for all } j \neq i\}$$

最大事後確率則 (2)

- ベイズの定理を用いれば，事後確率は

$$p(\omega_i \mid x) = \frac{p(x \mid \omega_i)\, p(\omega_i)}{p(x)}$$

- $p(x)$ はカテゴリによらないため，

$$p(\omega_i \mid x) \propto \underbrace{p(x \mid \omega_i)}_{\text{条件付き確率}}\, \underbrace{p(\omega_i)}_{\text{事前確率}}$$

- 最大事後確率則は，条件付き確率と事前確率の積を最大にするカテゴリに分類することと等価．

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**
- $p_e(\omega_i \to \omega_j)$：**カテゴリ** ω_i **に属するパターンが誤ってカテゴリ** ω_j **に分類される確率**
- **これはカテゴリ** ω_i **に属するパターンが決定領域** D_j **に入る確率と等価.**

$$p_e(\omega_i \to \omega_j) = \int_{x \in D_j} p(x \mid \omega_i) dx$$

最小誤識別率則 (2)

- $p_e(\omega_i)$ ：カテゴリ ω_i 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

$$
\begin{aligned}
p_e(\omega_i) &= \sum_{j \neq i} p_e(\omega_i \to \omega_j) \\
&= \sum_{j \neq i} \int_{x \in D_j} p(x \mid \omega_i) dx \\
&\quad + \int_{x \in D_i} p(x \mid \omega_i) dx - \int_{x \in D_i} p(x \mid \omega_i) dx \\
&= 1 - \underbrace{\int_{x \in D_i} p(x \mid \omega_i) dx}_{\text{正解率}}
\end{aligned}
$$

最小誤識別率則 (3)

- 全体の誤識別率 p_e ：
  $p_e(\omega_i)$ を全カテゴリーに対して平均したもの

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

- 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．

ベイズ決定則 (1)

- 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．
- 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．
- 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．
- **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

ベイズ決定則 (2)

- $l_{i,j}$ ：カテゴリ ω_i に属するパターンを誤ってカテゴリ ω_j に分類したときの**損失**(loss)

- **条件付き危険**(conditional risk) $R(\omega_j \mid x)$ ：パターン x をカテゴリ ω_j に分類したときの損失の期待値

$$R(\omega_j \mid x) = \sum_{i=1}^{m} l_{i,j}\, p(\omega_i \mid x)$$

ベイズ決定則 (3)

$$R(\omega_j \mid x) = \sum_{i=1}^{m} l_{i,j}\, p(\omega_i \mid x)$$

- 条件付き危険が最小になるカテゴリに分類.

$$x \in \omega_k : k = \arg\min_j R(\omega_j \mid x)$$

- 決定領域を次のように設定することと等価.

$$D_i = \{x \mid R(\omega_i \mid x) \le R(\omega_j \mid x) \text{ for all } j \ne i\}$$

ベイズ決定則 (4)

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

$$R = \int_D R(\omega_j \mid x)\, p(x)\, dx$$

但し，j 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

- 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．
- 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．
- ベイズ決定則を用いるのが最も自然．
- しかし，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4回)

1. 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．
2. ベイズ決定則で，損失が

$$l_{i,j} = \begin{cases} 0 & (i = j) \\ l & (i \neq j) \end{cases}$$

のとき，即ち，正しく識別した場合の損失が 0，誤って識別した場合の損失がカテゴリによらず一定値 l のとき，ベイズ決定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．

小レポート (続き)

3. 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？