パターン認識特論

Advanced Pattern Recognition

Toru Wakahara

授業のテーマと到達目標

統計的パターン認識のより進んだ理論的枠組みを体系的に学び，同時に具体的な適用法を学ぶことをテーマとし，

① 統計的パターン認識の生成モデルと識別モデルの代表的手法を理解して習得する。

② 手書き数字認識を例題に代表的認識手法を複数実装して認識実験を行い，各手法の得手不得手を知る。

を到達目標とする。

授業の概要

1. まず，ベイズの定理と最小誤り確率に基づく統計的パターン認識の原理を理解する。
2. 次に，生成モデルにおける混合ガウスモデルとEM解法，改良投影距離法，識別モデルとしてのニューラルネットワークとサポートベクターマシンを学ぶ。
3. 授業後半では，手書き数字認識を取り上げ，画像取得，前処理，特徴抽出，さらに授業前半で学んだ複数の認識手法の実装に取り組み，公開データを用いた認識実験と評価を行う。

授業計画

	内容
第1回	統計的パターン認識の考え方 --- 誤り確率最小化と最適な決定境界
第2回	生成モデルと識別モデル --- 確率密度関数の推定 vs. 識別関数の推定
第3回	生成モデルにおける確率密度関数の推定 --- 混合ガウスモデル法とそのEM解
第4回	生成モデルとしての改良投影距離法 --- 正規分布を仮定した高度な識別関数
第5回	識別モデルとしてのニューラルネットワーク --- 写像能力と誤差逆伝搬法，深層学習
第6回	識別モデルとしてのサポートベクターマシン --- カーネルトリックと解法
第7回	手書き数字認識系の構築（1） --- データ取得と前処理
第8回	手書き数字認識系の構築（2） --- 特徴抽出
第9回	手書き数字認識系の構築（3） --- 最近傍平均分類法およびk-NN分類法の実装
第10回	手書き数字認識系の構築（4） --- 改良投影距離法の実装
第11回	手書き数字認識系の構築（5） --- サポートベクターマシンの実装
第12回	手書き数字認識系の構築（6） --- 畳み込みニューラルネットワーク(CNN)の実装
第13回	手書き数字認識系の構築（7） --- 認識実験と評価
第14回	まとめ --- 最終レポート作成，成果発表

参考書:

[1] 石井健一郎, 上田修功, 前田英作, 村瀬洋著:
「わかりやすいパターン認識」, オーム社, 1998年.

[2] C.M. Bishop: “*Pattern Recognition and Machine Learning*”, Springer, 2006.

[3] 金谷健一著:「これなら分かる応用数学教室
－最小二乗法からウェーブレットまでー」,
共立出版, 2003年.

[4] 金谷健一著:「これなら分かる最適化数学
－基礎原理から計算手法までー」,
共立出版, 2005年.

[5] 斎藤康毅著:「ゼロから作るDeep Learning」,
オライリー・ジャパン, 2016年.

授業時間外の学習:

[1] 確率・統計，線形代数の復習

[2] 特殊関数の微積分の復習

[3] C/C++, MATLAB, Python, Java言語プログラミングの復習

講義ノート: 授業支援システムの教材にupload

patrec_master_1.pdf, …

講義の進め方:

1. パターン認識の理論的基礎を理解する
→ プログラミングで実装してみる

2. 手書き数字認識を例題に，認識系構築のための各段階の処理を実装する
→ 認識系の評価方法を学ぶ

3. 質問時間を十分に取る

評価方法:

宿題 40%　最終レポート 40%

平常点 20%

第１回講義「パターン認識特論」

[1] パターン認識とは何か

[2] 一つの例：文字認識

[3] 統計的パターン認識の考え方

[4] ベイズの定理と事後確率

[5] 決定領域とベイズ決定理論

パターン認識とは何か（1）

- 「パターン」とは何か

  ①型，類型

  ②図形，図像　（岩波国語辞典第六版）

- 「パターンとはそう見なすことで生じる」

  (*see* something *as* something)

  → 主観に強く依存する

  （渡辺慧："認識とパタン"，岩波新書）

パターン認識とは何か(2)

- 人間が主観的／概念的に把握している「パターン」を計算機で「認識」すること

“人工知能研究の夢”

- 工学的に実現すべきことは何か

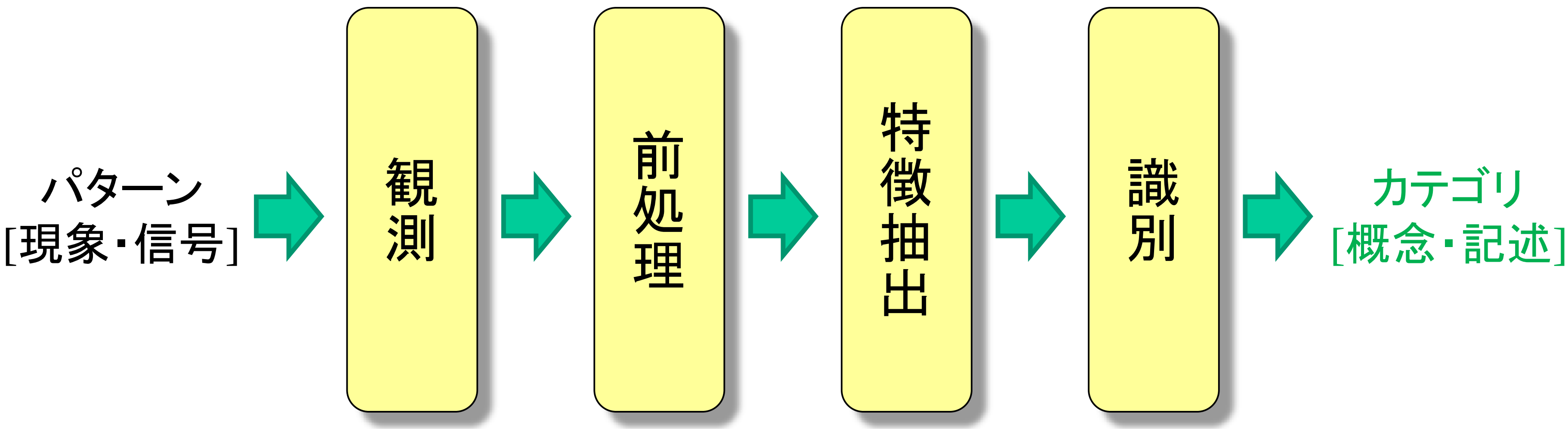
①「パターン」を数量的に定義すること

②「認識」する数学的手段を提供すること

①は発見法的，②は理論的

パターン認識とは何か（3）

■ パターン認識処理の汎用的なフレームワーク



- 前処理は，特徴抽出と識別に先立つために「前」と付き，雑音除去や平滑化，データ整形の処理から成る
- 知識処理により識別結果を修正する「後処理」が，識別の後段に加わることもある

パターン認識理論の枠組み

確率論的アプローチ *vs.* **決定論的アプローチ**

確率論的性質：
- パターン分布の推定
- 確率的な認識判断

決定論的振舞い：
- パターン生成規則
- パターン変形規則

最も汎用的かつ自然な枠組み
"統計的パターン認識理論"

- パターン認識研究の初期には決定論的アプローチも盛んだったが，確率論的アプローチが理論研究の大きな発展により主流となった
- 学習サンプル数が少ない場合は決定論的アプローチも有用である

統計的パターン認識の課題（1）

[1] 特徴空間をどう構成するか

→パターン認識の性能は特徴で決まるが，特徴抽出のガイドラインや理論研究は未だ確立していない

[2] 最適な特徴次元数はどう決めるのか

→多ければ多いほど良いわけでない
これについても理論研究は少なく，試行錯誤を伴う

[3] パターン認識は特徴から分類や回帰を行う非線形写像である

→どのような関数を用いたら所望の非線形写像を精度よく近似できるかが課題であるが，理論研究が最も進んでおり，現在も進行中である

統計的パターン認識の課題(2)

[4] 前処理の重要性

→パターン認識の負荷を軽減する効果が大きい
もっと重点的に研究すべきであろう

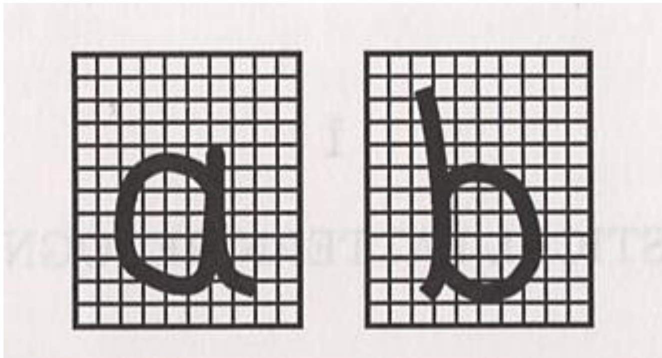
[5] 次元の呪いがある

→有限の学習サンプル数では高次元の特徴分布を
充分に推定できない
the small sample-size problemは理論研究が盛ん

[6] 汎化能力をどう評価するか

→未知サンプルに対する認識性能をどうしたら保証
できるかはパターン認識の究極の課題であり,
挑戦的で最先端の研究テーマである

一つの例：文字認識 ‘a’ と ‘b’ の識別



C_1 : ‘a’　　C_2 : ‘b’

$$\boldsymbol{x} = (x_1\ x_2\ \cdots\ x_d)^{\mathrm{T}}$$
$$x_i \in [0, 255]$$

8-bitの階調値を持つ
濃淡画像で与えられている

総画素数をdとすると可能な画像の数は
$$2^{8\times d} \cong 10^{0.3010\times 8\times d}$$

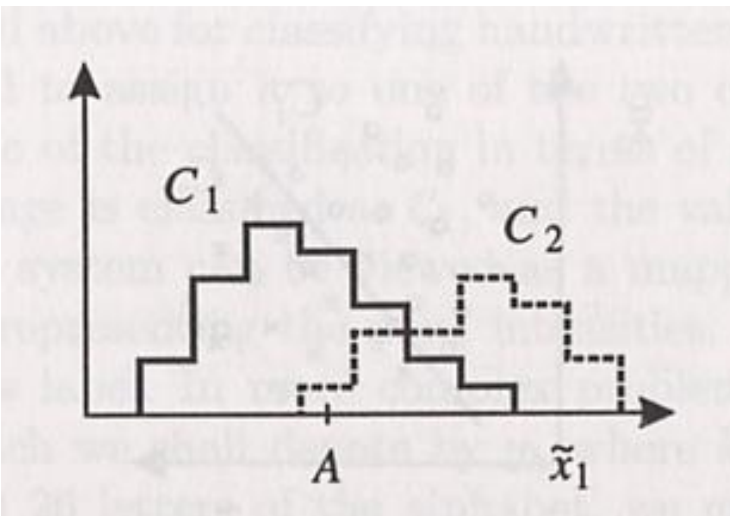
可能な全ての画像について分類結果を
人手でつけることは不可能である

パターン認識の課題：限られた学習サンプル集合から
未知サンプルを最適に識別する分類器をつくること

識別のための特徴抽出

特徴の抽出・選択:
① 対象の知識に基づく発見法的な手法
② 数学的な変数変換による次元圧縮

➢ 最適な特徴抽出・選択の理論は確立していない
試行錯誤で決めているのが現状である



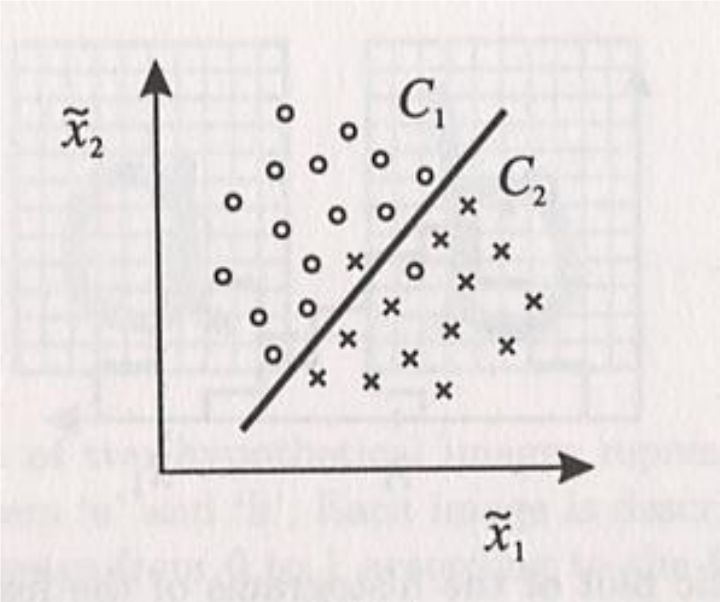
例えば,

$\tilde{x}_1$:文字の縦/横比

← ‘a’ と ‘b’ の学習サンプルでの縦/横比のヒストグラム

$\tilde{x}_1 = A$ ⇒ C_1, C_2 のどちらに属すると判断したらよいのか?

特徴増加の効果－Feature selection



① $\tilde{x}_2$の追加
→安定な決定境界

② $\tilde{x}_1, \tilde{x}_2$単独では
識別困難

では特徴増加で識別能力は単調に向上するか?
そうとも言えない → よく分からない

ただサンプル数が少ないと飽和する → 次元の呪い
特徴空間は与えられたものとして，誤り確率が最小となる
分類器を構築する → 統計的パターン認識の中心課題

パターン認識での統計的アプローチ

文字認識　→　誤り確率が最小となるように
入力文字のカテゴリ ‘a’ , ‘b’ を決定

➢ まず学習データを用いて，カテゴリの事前確率を求めておく

膨大数のサンプルを収集
→ ‘a’ と ‘b’ の出現頻度
ex. ‘a’ は ‘b’ の1.5倍出現

事前確率
$P(C_1) = 0.60$
$P(C_2) = 0.40$

➢ 入力文字を観測する前と後とでは，最適分類の戦略は変わるだろうか?

[1] 事前確率のみによる最適な決定は?
[2] 入力文字を観測してからの最適な決定は?

パターン認識のベイズ決定理論

$P(C_k)$: クラス C_k の事前確率

$P(X^l|C_k)$: クラス C_k 条件付きの X^l 出現確率

$P(C_k|X^l)$: X^l 条件付きのクラス C_k の事後確率

➢ 統計的パターン認識では，これらの確率を使い分けて，最適な分類を行う

未知パターンから X^l が観測されたとき，
事後確率 $P(C_k|X^l)$ が最大となるクラスに
決定するのが，誤り確率最小の最適な判断である

統計的パターン認識の基本原理

事後確率と統計的パターン認識

事前確率とクラス条件付き確率が分かっていると
事後確率を計算することができる！
ベイズの定理

- 事後確率を直接推定するより，事前確率とクラス条件付き確率を推定する方がはるかに容易であり，それにより事後確率も推定できる

↓

統計的パターン認識では，事前確率とクラス条件付き確率を正しく推論すれば，最適分類が実現できる

ベイズの定理 (Bayes' theorem)

$$P(C_k, X^l) = P(X^l | C_k) P(C_k)$$

$$P(C_k, X^l) = P(C_k | X^l) P(X^l)$$

$$P(C_k | X^l) = \frac{P(X^l | C_k) P(C_k)}{P(X^l)} \qquad (1)$$

事後確率

Bayes' theorem

➢ 統計的パターン認識の基本であり，この式は覚えること

ベイズの定理 (Bayes' theorem)

式(1)を $P(C_1|X^l) + P(C_2|X^l) = 1$ に代入すると

$$P(X^l) = P(X^l|C_1)P(C_1) + P(X^l|C_2)P(C_2)$$

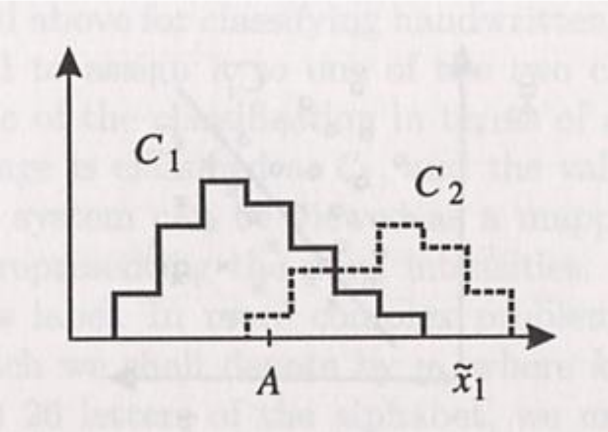
$$P(C_k|X^l) = \frac{P(X^l|C_k)P(C_k)}{P(X^l|C_1)P(C_1) + P(X^l|C_2)P(C_2)}$$

事後確率

Bayes' theorem

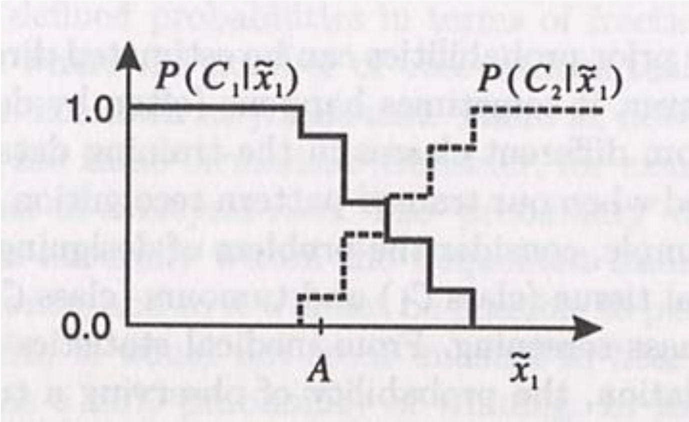
- 2クラス分類でのベイズの定理であり，この式も覚えること

推論と決定－Inference and decision



$P(\tilde{x}_1|C_1), P(\tilde{x}_1|C_2)$

$P(C_1) = 0.60$
$P(C_2) = 0.40$



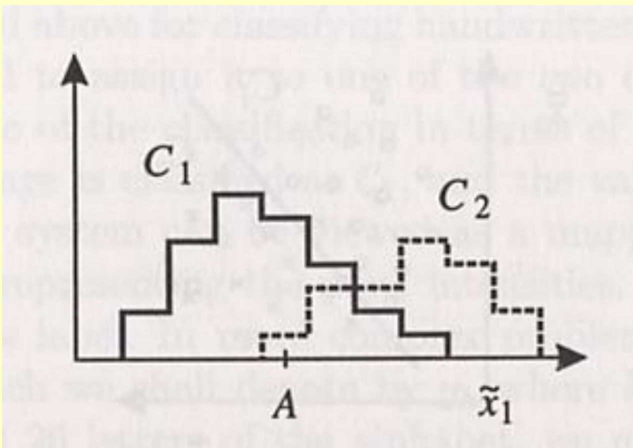
$P(C_1|\tilde{x}_1), P(C_2|\tilde{x}_1)$

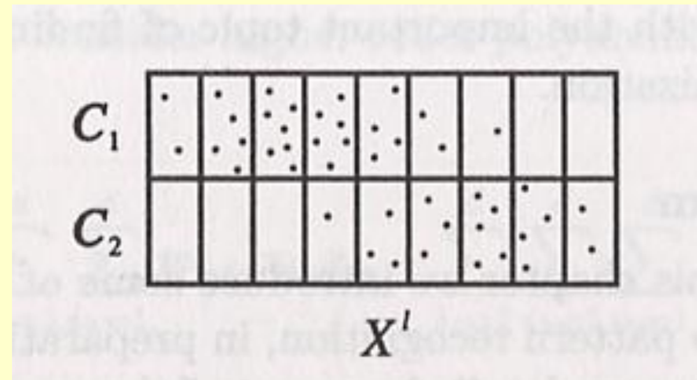
① 推論段階：学習データを用いて事前確率とクラス条件付確率を推定し，続いてベイズの定理より
→ 事後確率を推論する

② 決定段階：入力文字を観測したら，事後確率を計算して事後確率が最大となるクラスに分類すると
→ 誤り確率が最小となる

Quiz

事前確率とクラス条件付確率:





結合確率$P\left(C_k, X^l\right)$

右上図を用いて次の確率をそれぞれ既約分数で求めなさい．但し，$k = 1, 2$ および $l = 1, 2, \ldots, 9$.

(1) $P(C_k)$

(2) $P\left(X^l\right)$

(3) $P\left(C_k \middle| X^l\right)$

(4) $P\left(X^l \middle| C_k\right)$

- 右上図で，行方向に見ると $P(\cdot\,|C_k)$，列方向に見ると $P\left(\cdot\,\middle|X^l\right)$

確率から確率密度へ

離散変数 $\tilde{x}_1$ から連続変数 x への移行:

$$P(x \in [a, b]) = \int_a^b p(x)dx$$

$\boldsymbol{x} = (x_1 \ \cdots \ x_d)^{\mathrm{T}}$ の場合:

$$P(\boldsymbol{x} \in \mathcal{R}) = \int_{\mathcal{R}} p(\boldsymbol{x})d\boldsymbol{x}$$

期待値:

$$E[Q] \equiv \int Q(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} \cong \frac{1}{N}\sum_{n=1}^{N} Q(\boldsymbol{x}_n)$$

ベイズの定理 の一般形

$$P(C_k|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|C_k)P(C_k)}{\sum_{j=1}^{c} p(\boldsymbol{x}|C_j)P(C_j)} \qquad (2)$$

事後確率

尤度　　　事前確率

$$posterior = \frac{likelihood \times prior}{normalization\ factor}$$

事後確率

正規化因子

Bayes' theorem in general

最適な決定領域（1）

入力文字の $\boldsymbol{x}$ を観測して ‘a’, ‘b’ に分類する際誤り確率を最小とする最適な決定領域は?

$\mathcal{R}_1, \mathcal{R}_2$: クラス C_1, C_2 に分類する決定領域

$$P(\text{error}) = P(\,\boldsymbol{x} \in \mathcal{R}_2, C_1) + P(\,\boldsymbol{x} \in \mathcal{R}_1, C_2)$$

$$= P(\boldsymbol{x} \in \mathcal{R}_2 | C_1)P(C_1) + P(\boldsymbol{x} \in \mathcal{R}_1 | C_2)P(C_2)$$

$$= \int_{\mathcal{R}_2} p(\boldsymbol{x}|C_1)P(C_1)d\boldsymbol{x} + \int_{\mathcal{R}_1} p(\boldsymbol{x}|C_2)P(C_2)d\boldsymbol{x}$$

$$\rightarrow \min \ \ \text{for} \ \ \mathcal{R}_1, \mathcal{R}_2$$

最適な決定領域（2）

$$P(\text{error}) = P(\,\boldsymbol{x} \in \mathcal{R}_2, C_1) + P(\,\boldsymbol{x} \in \mathcal{R}_1, C_2)$$

$$= \int_{\mathcal{R}_2} p(\boldsymbol{x}|C_1)P(C_1)d\boldsymbol{x} + \int_{\mathcal{R}_1} p(\boldsymbol{x}|C_2)P(C_2)d\boldsymbol{x}$$

$$= \int_{\mathcal{R}_1+\mathcal{R}_2} p(\boldsymbol{x}|C_1)P(C_1)d\boldsymbol{x}$$

$$+ \int_{\mathcal{R}_1} [p(\boldsymbol{x}|C_2)P(C_2) - p(\boldsymbol{x}|C_1)P(C_1)]d\boldsymbol{x}$$

$$= P(C_1) + \int_{\mathcal{R}_1} [p(\boldsymbol{x}|C_2)P(C_2) - p(\boldsymbol{x}|C_1)P(C_1)]d\boldsymbol{x}$$

$$\rightarrow \min \ \ \text{for} \ \ \mathcal{R}_1, \mathcal{R}_2$$

最適な決定領域（3）

$$P(\mathrm{error}) = P(\,\boldsymbol{x} \in \mathcal{R}_2, C_1) + P(\,\boldsymbol{x} \in \mathcal{R}_1, C_2)$$

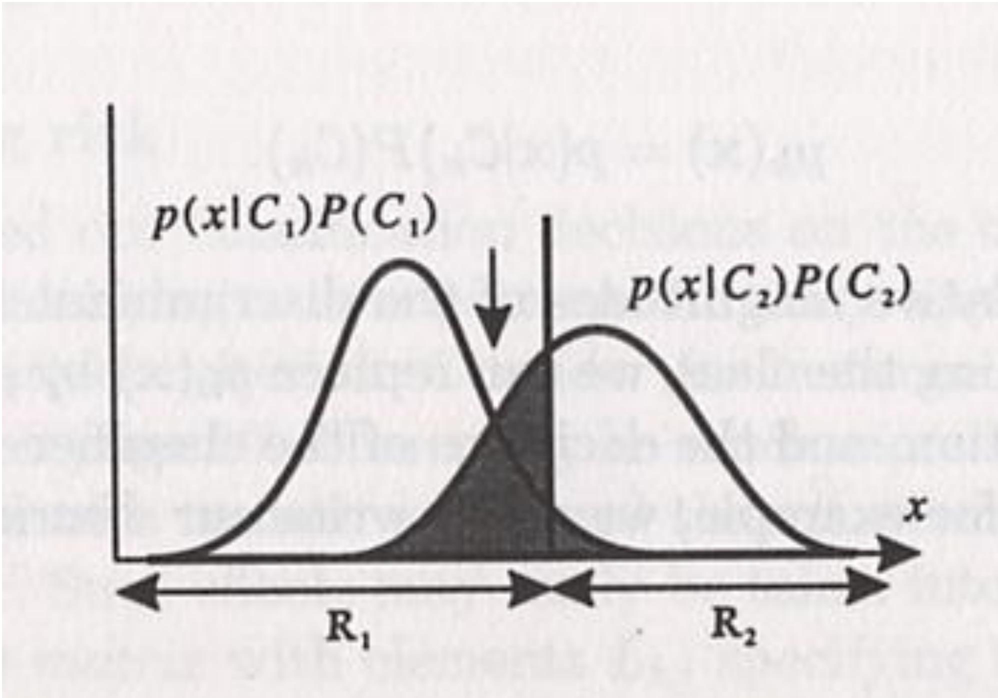
$$= P(C_1) + \int_{\mathcal{R}_1} [p(\boldsymbol{x}|C_2)P(C_2) - p(\boldsymbol{x}|C_1)P(C_1)]d\boldsymbol{x}$$

$$= P(C_1) + \int_{\mathcal{R}_1} [P(C_2|\boldsymbol{x}) - P(C_1|\boldsymbol{x})]p(\boldsymbol{x})d\boldsymbol{x}$$

$$\to \min \quad \mathrm{for} \quad \mathcal{R}_1, \mathcal{R}_2$$

$$P(C_1|\boldsymbol{x}) > P(C_2|\boldsymbol{x}) \to \boldsymbol{x} \in \mathcal{R}_1$$
$$P(C_2|\boldsymbol{x}) > P(C_1|\boldsymbol{x}) \to \boldsymbol{x} \in \mathcal{R}_2$$

決定領域とベイズ決定理論



$P(\mathrm{error})$は
影部分の面積

最小化する
$\mathcal{R}_1, \mathcal{R}_2$

$$p(\boldsymbol{x}|C_1)P(C_1) > p(\boldsymbol{x}|C_2)P(C_2) \rightarrow \boldsymbol{x} \in \mathcal{R}_1$$
$$p(\boldsymbol{x}|C_2)P(C_2) > p(\boldsymbol{x}|C_1)P(C_1) \rightarrow \boldsymbol{x} \in \mathcal{R}_2$$

多クラス分類での最適な決定領域

$$
\begin{aligned}
P(\text{correct}) &= \sum_{k=1}^{c} P(\boldsymbol{x} \in \mathcal{R}_k, C_k) \\
&= \sum_{k=1}^{c} P(\boldsymbol{x} \in \mathcal{R}_k | C_k) P(C_k) \\
&= \sum_{k=1}^{c} \int_{\mathcal{R}_k} p(\boldsymbol{x}|C_k) P(C_k) d\boldsymbol{x} \ \ \rightarrow \max \ \ \text{for} \ \{\mathcal{R}_k\}
\end{aligned}
$$

$$
p(\boldsymbol{x}|C_k)P(C_k) > p\left(\boldsymbol{x}\middle|C_j\right)P\left(C_j\right) \\
\text{for} \ \ \forall j \neq k \ \ \rightarrow \boldsymbol{x} \in \mathcal{R}_k
$$

リスクの導入

L_{kj}：本当は C_k に属するパターンを誤って C_j に分類した際の損失

R_k： C_k に属するパターンに対する期待損失

$$R_k = \sum_{j=1}^{c} L_{kj} \int_{\mathcal{R}_j} p(\boldsymbol{x}|C_k) d\boldsymbol{x}$$

$$R = \sum_{k=1}^{c} R_k P(C_k) = \sum_{j=1}^{c} \int_{\mathcal{R}_j} \left\{ \sum_{k=1}^{c} L_{kj} p(\boldsymbol{x}|C_k) P(C_k) \right\} d\boldsymbol{x}$$

リスク最小化での最適な決定領域

$$R = \sum_{j=1}^{c} \int_{\mathcal{R}_j} \left\{ \sum_{k=1}^{c} L_{kj}\, p(\boldsymbol{x}|C_k) P(C_k) \right\} d\boldsymbol{x}$$

$$\rightarrow \min \ \ \text{for} \ \{\mathcal{R}_j\}$$

$$\sum_{k=1}^{c} L_{kj}\, p(\boldsymbol{x}|C_k) P(C_k) < \sum_{k=1}^{c} L_{ki}\, p(\boldsymbol{x}|C_k) P(C_k)$$

$$\text{for} \ \forall i \neq j \ \rightarrow \boldsymbol{x} \in \mathcal{R}_j$$

課題１－１

Consider two non-negative numbers a and b, and show that, if $a \leq b$, then $a \leq (ab)^{1/2}$.
Use this result to show that, if the decision regions of two–class classification problem are chosen to minimize the probability of misclassification, this probability will satisfy

$$P(\text{error}) \leq \int_{\mathcal{R}_1+\mathcal{R}_2} \{p(\boldsymbol{x}|C_1)P(C_1)p(\boldsymbol{x}|C_2)P(C_2)\}^{1/2} d\boldsymbol{x}$$

Hint: Use the results obtained in pages 27-29.

課題1－2

Show that in a multiclass problem with c classes the probability of misclassification for the optimum classifier is bounded by

$$P(\mathrm{error}) \leq \frac{c-1}{c}$$

Hint: Show that for each $\boldsymbol{x}$ the maximum of $P(C_i|\boldsymbol{x})$, $i = 1,2,\ldots,c$, is greater than or equal to $1/c$, and

$$P(\mathrm{error}) = \int \big(1 - P(C_{max}|\boldsymbol{x})\big) p(\boldsymbol{x}) d\boldsymbol{x}\,,$$

where $P(C_{max}|\boldsymbol{x}) \equiv \max_{1 \leq i \leq c} P(C_i|\boldsymbol{x})$

課題１－１，１－２のレポート提出要領

A4用紙に，手書きもしくはMS Wordを用いて，解答する。

【タイトル】 課題１－１，１－２

提出年月日，学生証番号，氏名を必ず記すこと。

【提出期限】 次回（第2回）講義冒頭

☆質問は，W5023を直接訪れる（火，水，金）か，もしくは
若原（wakahara@hosei.ac.jp）までメールのこと。