

パターン認識と機械学習

Pattern Recognition and Machine Learning

Toru Wakahara

授業のテーマと到達目標

パターン認識と機械学習の基礎となる考え方とその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主要技法を理解し、特長を説明できる.
- ② 機械学習の代表的手法であるDeep Learningの基礎理論を理解し、Pythonによる具体的実装法を習得する.

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**の考え方を導入する。
2. 次に、**生成モデル**に基づくアプローチを説明し、中心的課題である**確率密度関数の代表的な推定法**を紹介する。
3. さらに、**識別モデル**に基づくアプローチを説明し、**線形識別関数法**から入り、より柔軟性のある**多層パーセプトロン**を紹介する。
4. 授業後半では、**Deep Learning**を取り上げ、理論を学びながら、**Pythonプログラミング**で実装を行う。

授業計画

	内容
第1回	統計的パターン認識の考え方 --- クラスと特徴の結合確率, ベイズの定理と事後確率
第2回	ベイズ決定理論 --- 誤り確率最小化, 最適な決定領域, 生成モデルと識別モデル
第3回	確率密度関数の推定(1) --- パラメトリックな方法としてのガウスモデルと最尤法
第4回	確率密度関数の推定(2) --- ノンパラメトリックな方法としてのカーネル法
第5回	識別モデルに基づく識別関数法 --- 線形識別関数による2クラスと多クラスの分類
第6回	ニューラルネットワーク(1) --- 単層パーセプトロンから多層パーセプトロンへ
第7回	ニューラルネットワーク(2) --- 多層パーセプトロンの写像能力と学習法
第8回	Deep Learning(1) --- Pythonプログラミング環境の設定とパーセプトロンの実装
第9回	Deep Learning(2) --- 各種活性化関数とニューラルネットワークの実装
第10回	Deep Learning(3) --- 各種損失関数を用いた学習アルゴリズムとその実装
第11回	Deep Learning(4) --- 誤差逆伝播法とその実装
第12回	Deep Learning(5) --- 学習に関する様々な技法の実装
第13回	Deep Learning(6) --- 畳み込みニューラルネットワークとその実装
第14回	まとめ --- 学習内容のまとめと重要ポイントの確認

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: “*Neural Networks for Pattern Recognition*,” Oxford University Press, 1995.
- [3] 杉山将著:「統計的機械学習ー生成モデルに基づくパターン認識」, オーム社, 2009年.
- [4] 金谷健一著:「これなら分かる応用数学教室ー最小二乗法からウェーブレットまでー」,
共立出版, 2003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.

講義の進め方:

1. パターン認識の理論的基礎を理解する
→ 計算問題を解く
2. Deep Learningの理論と実装をPythonで学ぶ
→ プログラミングを通して体得する
3. 講義ノート: 授業支援システムの教材にupload
PatrecML_1.pdf, ...

評価方法:

平常点 20点 宿題 20点

定期試験 60点

第1回講義「パターン認識と機械学習」

統計的パターン認識の考え方

[1] パターン認識とは

[2] 一つの例: 文字認識

[3] 事前確率とクラス条件付き確率

[4] ベイズの定理と事後確率

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, 1978

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”

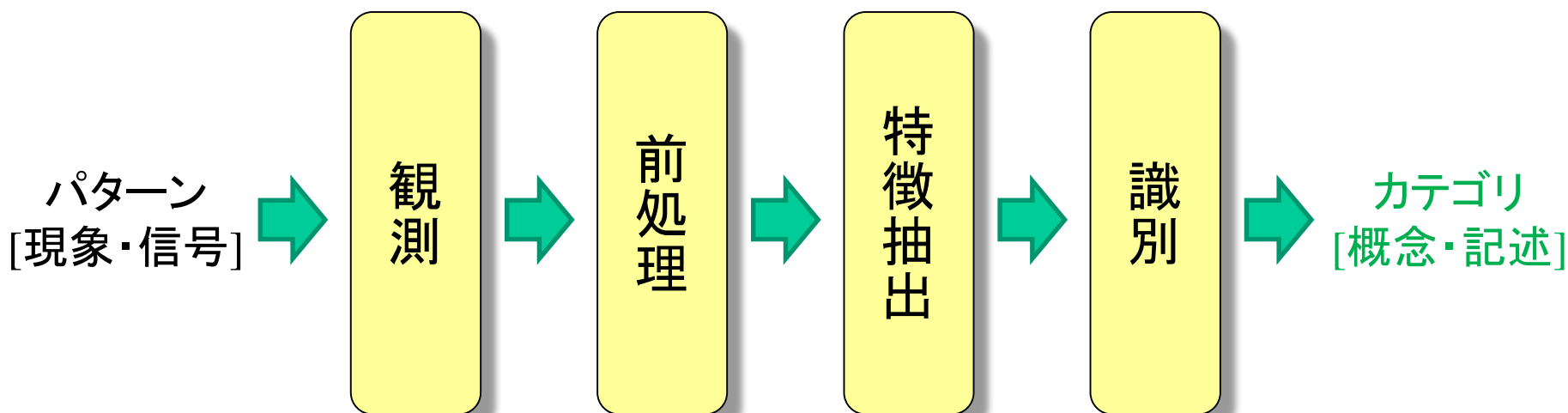


- 工学的に実現すべきことは何か
 - ① 「パターン」を数量的に定義すること
 - ② 「認識」する数学的手段を提供すること

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.

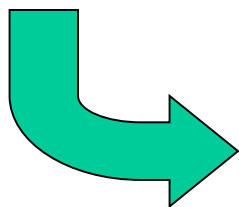
決定論的アプローチ

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:

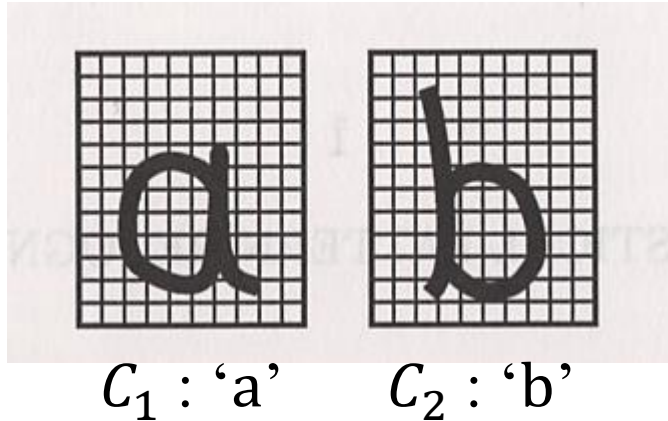
パターン生成規則
パターン変形規則



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

一つの例：文字認識‘a’と‘b’の識別

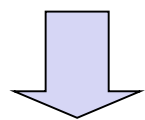


$$\mathbf{x} = (x_1 \ x_2 \ \cdots \ x_d)^T$$

$$x_i \in [0, 255]$$

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

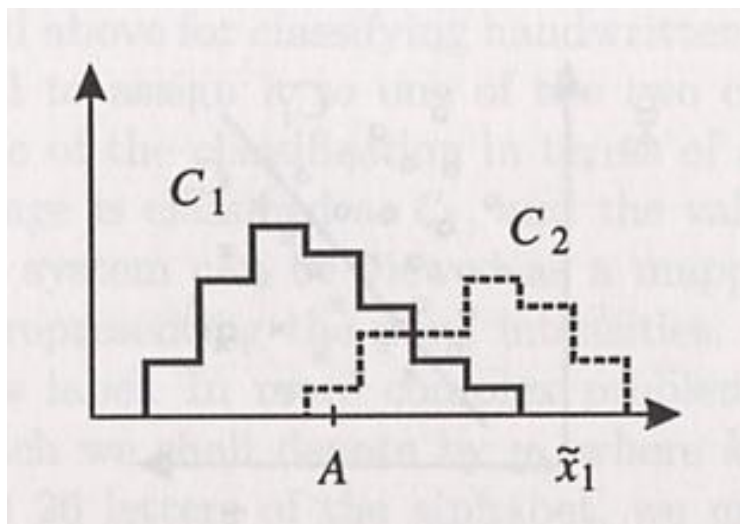
パターン認識の課題：限られた学習サンプル集合から
未知サンプルを最適に識別する分類器をつくること

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

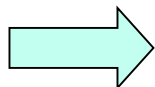


例えば,

\tilde{x}_1 : 文字の縦/横比

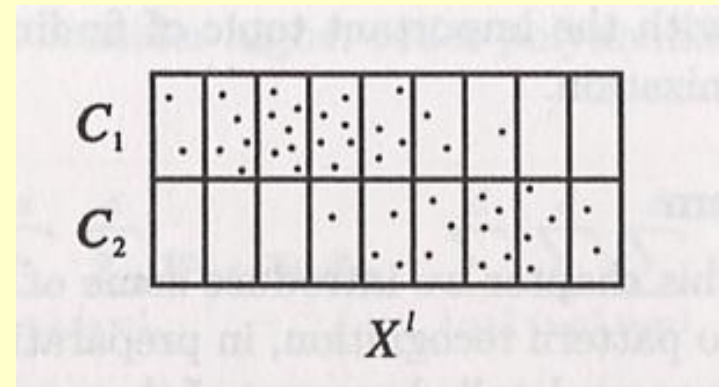
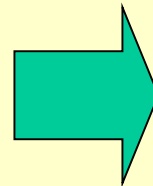
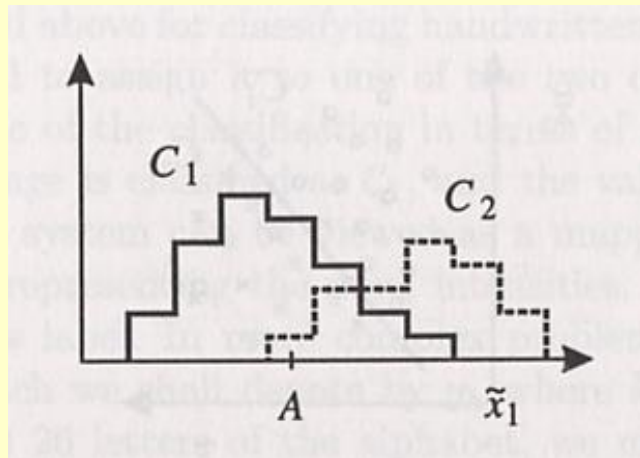
← 'a' と 'b' の学習サンプルでの
縦/横比のヒストグラム

$$\tilde{x}_1 = A$$



C_1, C_2 のどちらに属すると
判断したらよいのか?

クラスと特徴の確率的振舞い



(C_k, X^l) 分布の行列表現

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 $l = 1, 2, \dots, 9$ 。

事前確率: $P(C_k)$

クラス条件付き確率: $P(X^l | C_k)$

- ▶ クラス条件付き確率は, 右上図で, 行方向に見て計算すればよい。

Quiz

前頁で計算した解答を下表に埋めなさい。

事前確率:

$P(C_1)$	$P(C_2)$

クラス条件付き確率:

$P(X^1 C_1)$	$P(X^2 C_1)$	$P(X^3 C_1)$	$P(X^4 C_1)$	$P(X^5 C_1)$	$P(X^6 C_1)$	$P(X^7 C_1)$	$P(X^8 C_1)$	$P(X^9 C_1)$

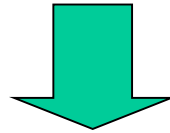
$P(X^1 C_2)$	$P(X^2 C_2)$	$P(X^3 C_2)$	$P(X^4 C_2)$	$P(X^5 C_2)$	$P(X^6 C_2)$	$P(X^7 C_2)$	$P(X^8 C_2)$	$P(X^9 C_2)$

ベイズの定理(1)

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

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

product rule of probability



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

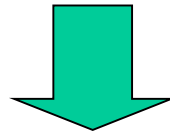
事後確率

➤ 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(2)

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

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



$$P(C_k|X^l) = \frac{P(X^l|C_k)P(C_k)}{P(X^l|C_1)P(C_1) + P(X^l|C_2)P(C_2)} \quad (2)$$

事後確率

➤ 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

ベイズの定理(3)

クラス数が c 個の場合の一般式は次の通り.

$$P(C_k|X^l) = \frac{P(X^l|C_k)P(C_k)}{P(X^l)} = \frac{P(X^l|C_k)P(C_k)}{\sum_{j=1}^c P(X^l|C_j)P(C_j)} \quad (3)$$

事後確率

注. 総確率の法則 (sum rule of probability) から

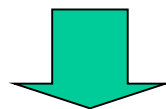
$$P(X^l) = \sum_{j=1}^c P(C_j, X^l) = \sum_{j=1}^c P(X^l|C_j)P(C_j)$$

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確率から事後確率をできるだけ正しく推論すれば、最適分類を実現できる

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

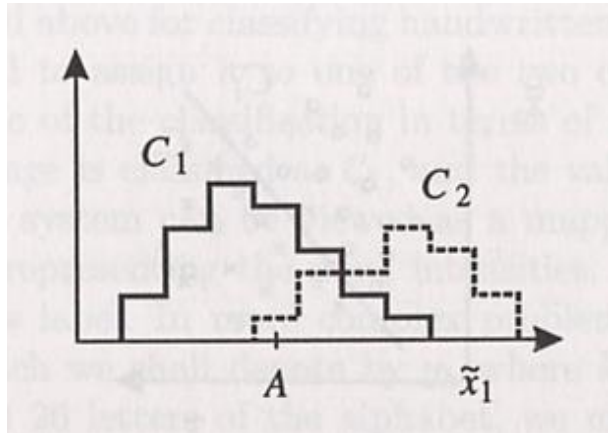
$P(C_k | X^l)$: X^l 条件付きのクラス C_k の事後確率

▶ 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則

推論と決定 — Inference and decision

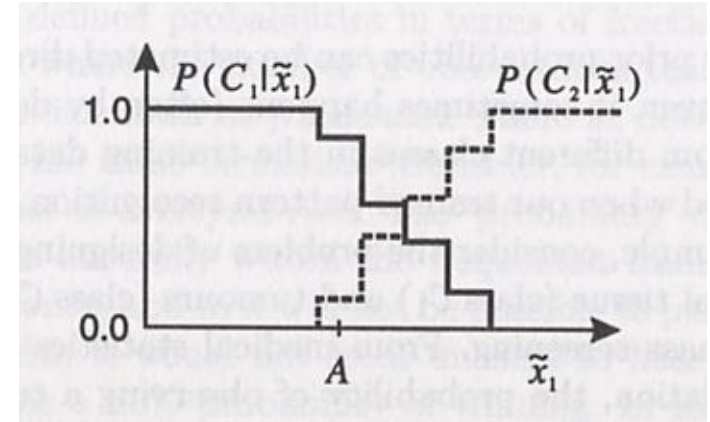


$$P(\tilde{x}_1|C_1), P(\tilde{x}_1|C_2)$$



$$P(C_1) = 3/5$$

$$P(C_2) = 2/5$$



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

- 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する. 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, ①で推論された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

第1回宿題

【問題1】

今回の講義ノート14,15頁で計算して求めた事前確率とクラス条件付き確率:

$$P(C_k) \quad (k = 1, 2)$$

$$P(X^l | C_k) \quad (k = 1, 2, l = 1, 2, \dots, 9)$$

を用いて, 講義ノート17頁にある式(2)から事後確率:

$$P(C_k | X^l) \quad (k = 1, 2, l = 1, 2, \dots, 9)$$

を求めなさい.

ただし, 計算過程を省略せずに記すこと.

第1回宿題(続き)

【問題2】

2つの箱がある. 第1の箱には, リンゴが7個, オレンジが5個入っている. 第2の箱には, リンゴが6個, オレンジが8個入っている. このとき, 第1の箱を確率 $1/3$, 第2の箱を確率 $2/3$ で選択して, 中から果物を1つ取り出すものとする.

今, 箱を選んで果物を取り出したら, リンゴであった. 選んだ箱が第1の箱である確率を計算しなさい.

ヒント: ベイズ則を用いなさい. 原因 x が箱の選択, 結果 y が取り出した果物の種類, に対応するとして, $P(x|y)$, すなわち, $P(\text{第1の箱}|\text{リンゴ})$ を計算しなさい.

宿題の提出要領

【提出方法】

- ・A4用紙に手書きもしくはMS Wordで作成
- ・複数枚の場合はホッチキス止め
- ・第1頁の冒頭に

「パターン認識と機械学習」 第〇回宿題

提出年月日 2019年△月□日

学生証番号 氏名

を必ず記す

【提出期限】

- ・次回講義の冒頭, 遅刻は認めず