

まえがき

近年、「ビッグデータ」や「ビジネスアナリティクス」といった言葉がニュースや雑誌記事を賑わすようになり、大規模データから必要な情報を抽出したり、データを自動分類したりする技術の重要性がますます高まっている。パターン認識と統計的学習の理論はその中心的な役割を持つ技術であるが、従来これらの理論は音声認識や画像認識、文字認識といった問題の中心的課題であり、情報通信工学、情報工学、電気工学、電子工学、人工知能といった分野で扱われてきたものである。これらの応用例では、例えば葉書の郵便番号の自動読み取りや、デジタルカメラの人物（顔）検出機能のような、機械やコンピュータが自動的に情報を認識するための固有技術の側面が強かった。最近になって、インターネット上でECサイトが本格的に普及したことにより、ユーザへのレコメンデーション（情報推薦）のような新たなウェブマーケティングツールが登場し、これらのツールの技術を支える理論として、パターン認識や統計的学習の理論が重要な位置を占めるようになってきた。データサイエンティストを抱える多くの企業では、日々蓄積される大規模データを分析し、なんらかの仮説や結論を導くためのツールとして、パターン認識と統計的学習の理論として長年にわたって発展してきた方法論を活用し始めている。すなわち、これらの理論は、経営学や経営工学、マーケティングといった分野に応用され、融合し、さまざまな方向へ発展していくステージに入ったと考えられる。

本書は、このような最近の応用の方向性も考慮しつつ、初学者が一通りのパターン認識と統計的学習の基礎を学ぶことができるように、基本的な内容に絞って書き下ろしたものである。大学3年生向けの授業で毎年配布して使用していた講義資料をベースとし、基礎を逸脱しない範囲で比較的最近のトピックや多少高度な内容も追加し、全体を書き直した。そのため、内容的に他書と書きぶりの異なる点や、一般的ではない観点が多少あるかもしれない。それらの点については、ご容赦いただければ幸いである。パターン認識や学習理論の教科書としては、すでに数多くの名

著がある^{1)~13)}†。どれも特徴を持った良著であるので、本格的にパターン認識を学習したい読者は、これらも参考にすることを勧める。

本書では、大学生や大学院生が実際にパターン認識の方法を実装し、実際のデータを分析しながら自ら手法を改良できるように、C言語によるプログラムを以下のウェブページで公開している。卒業論文や修士論文のためにプログラムが必要な読者は、利用していただきたい。

<https://www.coronasha.co.jp/static/02479/>

(ユーザ ID : , パスワード :)

本書の記載内容に誤りがあった場合も、このウェブページを通じて正誤表を公開したい。もし、内容などの誤りに気づいたら、コロナ社を通じて著者までご連絡いただければ幸いである。

最後に、本書をまとめる機会を与えていただいたコロナ社の皆様に深く感謝する。また、日頃よりご指導をいただいている早稲田大学理工学術院名誉教授の平澤茂一先生、同大学教授の松嶋敏泰先生、教学も含めさまざまな面でご協力をいただいている早稲田大学創造理工学部経営システム工学科、ならびに湘南工科大学工学部情報工学科の皆様にも深く感謝したい。特に、いつも著者らの研究活動に多大なサポートをしていただいている三川健太助手には深く感謝する。また、湘南工科大学の鈴木 誠先生、早稲田大学メディアネットワークセンターの石田 崇先生、須子統太先生、堀井俊佑先生には、共同研究者として刺激をいただいている。さらに、著者らの研究活動には、後藤研究室、小林研究室の大学院生、学部生のメンバー諸氏が多大な貢献をしてくれている。これらの多大なご協力をいただいているすべての方々に感謝の意を表す。

最後に、日頃から好き勝手な研究教育活動に多くの時間を割いている著者らを温かく見守っていただいている美穂さん、昌子さんに心より感謝する。

2014年3月

著者

† 肩付き番号は巻末の引用・参考文献を示す。

目 次

1. パターン認識と統計的学習の概要

1.1	パターン認識とは	2
1.1.1	パターン認識問題の例	2
1.1.2	パターン認識問題の基本モデル	3
1.1.3	特徴空間と特徴ベクトル	5
1.2	ベイズ識別規則と識別関数法	6
1.2.1	ベイズ決定による最適分類	6
1.2.2	識別関数による解釈	8
1.2.3	正規分布を仮定した場合の展開	8
1.2.4	線形識別関数によるパターン分類	11
1.3	統計的学習とパターン認識	13
	章 末 問 題	16

2. 特徴空間の構成と統計的性質

2.1	特徴空間と統計量	18
2.1.1	特徴空間と特徴ベクトル	18
2.1.2	パターン分布を測る尺度	19
2.1.3	データの基準化	20
2.1.4	特徴の選択	21
2.2	特徴空間の評価	22
2.2.1	パターン分布の評価	23
2.2.2	ベイズ誤り率	24
2.2.3	特徴空間と次元の呪い	25
	章 末 問 題	28

3. 線形識別の方法

3.1 線形識別器とパーセプトロン	29
3.2 フィッシャーの線形判別法	31
3.3 誤差評価関数による線形識別器の獲得	35
3.3.1 2乗誤差評価に基づくウエイト推定	36
3.3.2 Widrow-Hoff の学習規則	38
章 末 問 題	40

4. ナイーブベイズ法

4.1 ナイーブベイズ法の概要	41
4.2 高次元かつスパースな問題に対する対応	43
4.3 改良ナイーブベイズ法	44
4.4 ナイーブベイズ法の解釈	45
章 末 問 題	46

5. 線形部分空間による次元縮約

5.1 主成分分析	47
5.1.1 正規直交軸への射影	48
5.1.2 分散最大化による定式化	50
5.1.3 2乗誤差最小化による定式化	52
5.1.4 主成分得点	54
5.2 主成分分析と特異値分解	55
5.2.1 特異値分解	55
5.2.2 主成分分析との関係	58
章 末 問 題	59

6. テンプレートマッチングと k 最近傍識別法

6.1 特徴パターンとの照合によるパターン判別	60
6.1.1 最小距離に基づく判別	60

6.1.2	類似度に基づく判別	62
6.1.3	代表ベクトルの決定	63
6.2	k 最近傍識別法	64
6.2.1	k 最近傍識別法による分類	64
6.2.2	最近傍決定とベイズ決定の誤り率	65
6.3	メトリックラーニング	69
	章末問題	71

7. 決定木

7.1	分類木と回帰木	72
7.2	分類木の学習法	74
7.3	回帰木の学習法	81
7.4	学習アルゴリズムの拡張	82
	章末問題	84

8. 集団学習法

8.1	バギング	85
8.2	ブースティングと AdaBoost	90
8.3	ランダムフォレスト	98
8.4	ランダムフォレストによる特徴の重要度算出	104
8.5	定理8.1の証明	106
	章末問題	107

9. 非線形判別関数とニューラルネットワーク

9.1	区分的線形識別関数	109
9.2	階層型ニューラルネットワーク	111
9.2.1	生体の情報処理モデル	111
9.2.2	ニューラルネットワークのモデル	112
9.2.3	ニューラルネットワークの学習	114
	章末問題	120

10. カーネル法

10.1 リッジ回帰	121
10.2 特徴ベクトルとカーネル	124
章末問題	133

11. サポートベクトルマシン

11.1 サポートベクトルマシン	134
11.2 カーネルを用いたサポートベクトルマシン	141
11.3 ソフトマージンサポートベクトルマシン	144
11.4 サポートベクトルマシンの効率的学習法	150
章末問題	154

12. 関連ベクトルマシン

12.1 関連ベクトルマシンの概要	155
12.2 RVM の効率的学習法	161
12.3 RVM の予測法	166
章末問題	168

13. 二値判別器の組合せによる多値分類法

13.1 基本的な方法	169
13.1.1 one-vs-the rest 分類法	169
13.1.2 one-vs-one 分類法	171
13.2 ECOC 法	172
13.2.1 Exhaustive Code による構成法	172
13.2.2 Reed-Muller 符号による構成法	173
13.2.3 その他の方法	175
13.2.4 分類法	175
章末問題	177

14. 学習モデルと統計的推定

14.1 確率モデルとベイズ識別	178
14.2 パラメトリックな手法と統計的推定	180
14.2.1 パラメータの推定問題	180
14.2.2 ベイズ推定	184
14.2.3 尤度・事後確率の最大化	186
14.2.4 予測問題	187
14.2.5 ベルヌーイ試行のベイズ推定	189
章末問題	192

15. 潜在クラスモデル

15.1 混合正規モデル	194
15.2 EM アルゴリズム	198
15.3 潜在クラスモデルの例	200
15.3.1 多次元混合正規モデル	200
15.3.2 混合 Polya 分布モデル	201
15.3.3 PLSI と LDA	202
章末問題	204

16. 統計的モデル選択とモデル平均化法

16.1 多項式回帰モデルの例	205
16.2 階層モデル族	206
16.3 統計的モデル選択問題	209
16.4 モデル選択基準	211
16.4.1 赤池の AIC	212
16.4.2 Schwarz の BIC	215
16.4.3 Rissanen の MDL 基準	216
16.5 一致性の議論	217
16.5.1 最適モデルと一致性	217

16.5.2	モデルの事後確率の漸近的性質	219
16.5.3	情報量基準と一致性	220
16.5.4	モデル選択基準に関するさまざまな議論	221
16.6	モデル平均化法	222
章 末 問 題		224
付 録		226
A.1	ベクトル空間と関数の最適化	226
A.1.1	多次元2次関数	226
A.1.2	一般の関数	227
A.2	ラグランジュの未定乗数法	227
A.2.1	例 題	227
A.2.2	問題の一般定式化	229
A.2.3	ラグランジュの未定乗数法	229
A.2.4	ラグランジュの未定乗数法の解釈	231
A.3	固有値と固有ベクトル	232
A.3.1	線形変換	232
A.3.2	固有値と固有ベクトル	233
A.3.3	行列の基本的性質	233
A.4	多次元正規分布	234
A.4.1	2次元正規分布：無相関の場合	234
A.4.2	2次元正規分布：相関がある場合	235
A.4.3	多次元正規分布	237
引用・参考文献		238
索 引		242

1

パターン認識と統計的学習の概要

人間は五感を使って身のまわりの環境をたえずモニタリングし、瞬時にその置かれた状況を把握することができる。「なにかが変だぞ」と周囲の異常に気づいたり、海外旅行中に「あそこに日本人がいる」と発見したりすることもある。そもそも、同じように二足で歩く多くの人の中からは、自分と同じ日本人を見つけられるのはなぜだろうか。逆に、日本人が欧米人を見るとき、どこの国の人かを瞬時に見分けるのは難しいだろう。フランス人をイタリア人だと紹介されても、多くの日本人はその間違いを指摘できないかもしれない。しかし、中国人と日本人の見分けであれば、少なくとも欧米人よりはうまく見分けられそうである。

このように、人間は、瞬時に五感から得られた情報を処理して違いを見分けたり、それがなんであるかを認識したりすることができる。人間が行うこのような高度な情報処理の一つがパターン認識 (pattern recognition) である。パターン認識とは、観測された情報にあらかじめ与えられた概念を付与する処理のことをいう。一般に概念は、クラス (class) やカテゴリ (category) によって表現され、パターン認識とは「入力されたパターン情報に、それが属するカテゴリやクラスの情報を付与する処理」ということができる。例えば、空を飛ぶ生物を見て「あれは鳥だ」と認識することは、その飛ぶ姿や形状から、その対象を「鳥」という概念に対応付けることである。人間は非常に高度なパターン認識能力を持っており、「怪しい人」をなんとなく見分けたり、手書きで乱雑に書かれた文字を読んだりすることができる。

このようなパターン認識の処理をコンピュータで実現できれば、これは非常に有効な技術となりうる。しかし、先に述べたように、日本人は日本人と他のアジア人の見分けをある程度行えるが、欧米人の見分けは難しいというように、このパターン認識の能力には「それまでの経験値」が非常に密接に関わっているように思われる。したがって、パターン認識の精度は、人間世界においても経験や学習と密接な

関わりがあることが理解できるであろう。これは機械によるパターン認識においても同様であり、パターン認識と学習は切り離して議論することはできない。

コンピュータなどの機械に高度な情報処理を担わせようとする**知識情報処理** (intelligent information processing) の分野では、過去の事例によって、分類規則や知識を更新することを**学習** (learning)、または**統計的学習** (statistical learning) という。情報処理の分野においても、パターン認識は学習という概念と密接な関わりを持つ。ここでは、パターン認識の問題と統計的学習について、その概略を述べる。

1.1 パターン認識とは

1.1.1 パターン認識問題の例

人間が行う基本的な情報処理はパターン認識であるともいえる。ここでは、人間が行う高度なパターン認識の例を示すことにより、扱う問題の具体的なイメージを与えてみよう。

- 葉書の郵便番号を読み取り、郵便番号によって宛先を自動識別することができる。葉書の処理プロセスが飛躍的に効率化される。葉書に書かれた郵便番号は、送り主による手書きのものが多く、ばらつきのある文字を通常は人手によって正しく読み取らなければならない。すなわち、人手で郵便番号の升目内に書かれた記号情報を読み取り、これを番号に対応付ける必要がある。このような文字を認識する問題は**文字認識**、特に手書き文字の認識はワープロ印刷の文字の認識よりもはるかに難しいことから、**手書き文字認識**と呼ばれる。
- 人間は、友人を見かけて、瞬時にだれであるかを理解することができる。これは、友人の顔や声、体型などのパターン情報を過去の経験と照らし合わせ、その人がだれであるのかを特定する処理である。五感から得られるパターン情報に、個人名を付与する問題と考えることができる。
- 医者は、患者を診察して、精密検査が必要であるか否かを判断することができる。この高度な判断もパターン認識である。患者の症状や簡単な診察結果から、なんらかの重大な疾患の可能性を判断し、さらなる精密な検査の要否を付与する問題と考えることができる。

- 気象予報士は、気象図などの情報から、今後の天気の動向について予測を立てることができる。この問題も、現在の気象情報に基づいて、未来の天候を付与するパターン認識の問題と捉えることができる。
- 人間は、人が発した声から、その音の意味する文字を瞬時に聞き分けることができる。これは、周波数という情報を耳でキャッチし、その音に文字や文章を付与するパターン認識の問題である。このような音を認識する問題を音声認識という。

以上のように、パターン認識の問題はあらゆる場面で見かける。高度なパターン認識を実行できる機械を製造することができれば、それは指紋同定や DNA 鑑定、話者同定など、さまざまな最先端機器として応用が可能であり、工学的にたいへん有用である。このような、パターン認識機械の設計や製造は、長年多くの研究者によって取り組まれてきた。その成果として、指紋認証や文字認識、顔認識など、さまざまなパターン認識技術が実用化され、商品化のレベルに達している。しかしながら、それらの実用化に非常に長い研究時間を要していることからわかるように、パターン認識の自動化は、人間が予想する以上に難しい問題である。

1.1.2 パターン認識問題の基本モデル

音声認識や動画像のパターン認識など、時系列的な変化を取り扱う必要がある問題も多く存在するが、ここでは基本的なパターン認識のモデルを理解するために、時間変化を考慮しない静的なパターン認識問題について解説する。

パターン認識は、ある観測されたパターン情報を p とし、この情報からそのパターンが属するクラスのラベル c を求めることである。クラス (class) はカテゴリ (category) とも呼ばれる。すなわち、パターン認識の問題は、センサによって観測されたパターン情報 p に対し、それが属するクラス c を対応付ける写像 $f: p \rightarrow c$ を求めることに帰着する。

典型的なパターン認識システムは、まず、一般に膨大な情報であるセンサ入力 p から不必要な情報を削ぎ落とし、パターンの識別に有用な形の情報へと変換する。その際、クラス c による変動との独立性が高く、比較的容易に除去できる情報が多々存在したり、情報のスケールがまちまちであることも多い。前処理 (preprocessing) によって不必要な情報を削除したり、データを正規化 (normalization) することで、

必要な情報を保持しつつ、後のデータの取り扱いを容易にすることができる。その後、対象とするパターン認識の問題に重要な情報を抽出する処理が行われる。これを**特徴抽出**と呼ぶ。その結果、得られる量を**特徴量** (feature) という。各パターン情報は、この複数の特徴量からなる**特徴ベクトル** (feature vector) によって表現することができる。図 1.1 に特徴ベクトルの生成例を示す。

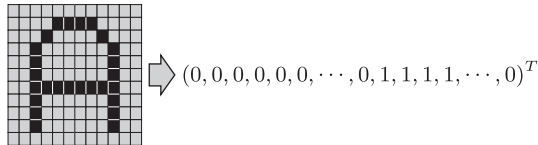


図 1.1 特徴ベクトルの生成例

これらの特徴ベクトルから最終的にそれが属するクラスを推定する処理を、**分類** (classification), または**識別** (discrimination) という[†]。特徴量からなる特徴ベクトルを用いて、どのようにクラスに分類するかには、さまざまな方法が考えられる。近年、多くの応用において成功を取めているのは、訓練データ (学習データ) を用い、統計的学習に基づいて、特徴ベクトルからクラスを帰納的に与える関数を推定する方法である。あらかじめ学習用の特徴ベクトルとクラスのセットを複数用意しておき、これらの正解データからその背後に存在する関係を推定し、新たな入力データの分類に用いる。ここで、このような学習のためには、パターンを分類するためのクラスがあらかじめ決まっている、正解データが複数与えられる必要があることに注意しよう。このような特徴ベクトルのクラスへの対応付けの問題を**パターン分類問題**という。また、得られた分類のルールや、分類を行う機械は、**分類器**または**判別器**と呼ばれる。

これに対し、問題によっては、パターンをどのようなクラスに分類すべきかもわからず、正解ラベルをあらかじめ与えようがないこともある。このような複数の特徴ベクトルが与えられ、これらをなんらかの基準で似たもの同士にグルーピングし

[†] 判別ともいう。パターン分類問題はさまざまは分野で取り扱われてきたため、「分類」や「識別」、「判別」といった語が、それぞれの文脈で混ざって使われているのが現状である。例えば、統計学の分野には「判別分析」という手法があるが、「分類分析」とは呼ばない。あとで出てくる「識別関数」も「分類関数」とは呼ばない。逆に、クラス数が多いとき、「多値識別」よりも「多値分類」が使われることが多い。

て、いくつかのクラスに分類することを、**クラスタリング (clustering)**、または**クラスタ分析 (cluster analysis)** という。

正解データが与えられ、これを学習してパターン分類器を構成するパターン分類問題の学習は**教師付き学習 (supervised learning)** と呼ばれる。一方、クラスタリングのように、正解のクラスラベルが与えられずに、自己組織化のように学習する方法を**教師なし学習 (unsupervised learning)** という。

1.1.3 特徴空間と特徴ベクトル

前項でも述べたように、与えられたパターンからその特徴を求める処理を**特徴抽出**という。得られた特徴は一般に

$$\mathbf{x} = (x_1, x_2, \dots, x_d)^T \quad (1.1)$$

のような**特徴ベクトル**として表現され¹⁾、 d 次元の特徴パターン空間上の1点で表される。ただし、 T は転置を表す。 d は一つのパターンから得られる特徴の数で、特徴パターン空間の次元を意味している。 \mathbf{x} が d 次元連続ベクトルであれば $\mathbf{x} \in \mathcal{R}^d$ であり、 \mathbf{x} の要素が $\{0, 1\}$ である0-1ベクトルであれば $\mathbf{x} \in \{0, 1\}^d$ のようになる²⁾。このようにして構成される特徴空間に対して、 \mathbf{x} のユークリッドノルム $\|\mathbf{x}\|$ を

$$\|\mathbf{x}\| = \sqrt{\|\mathbf{x}^T \mathbf{x}\|} = \sqrt{(x_1)^2 + (x_2)^2 + \dots + (x_d)^2} \quad (1.2)$$

で定義しておく。いま、特徴ベクトル \mathbf{x} で表されるパターンが属するクラスを c と記述しよう。クラスの集合を $\mathcal{C} = \{c_1, c_2, \dots, c_M\}$ 、その大きさを $M = |\mathcal{C}|$ とする。パターン認識では、得られた特徴ベクトルから、そのパターンが属するクラスを推測することが主たる問題となる。そのために、しばしば**確率モデル**が登場する。本書では、 \mathbf{x} の**確率**、あるいは**確率密度**を $p(\mathbf{x})$ のように記述する³⁾。

¹⁾ 慣例的な問題であるが、ベクトルを縦ベクトルで定義する教科書も多いので、本書においてもベクトルは縦ベクトルで定義する。しかし、実際に縦ベクトルで表記すると、より多くの紙面が必要になるため、横ベクトルを転置する形式で表記している。

²⁾ \mathcal{R}^d は d 次元のユークリッド空間を意味する。また、 $\{0, 1\}^d$ は、 $\{0, 1\}$ の d 回の直積 $\{0, 1\} \times \{0, 1\} \times \dots \times \{0, 1\}$ を表す。

³⁾ $p(\mathbf{x})$ は、 \mathbf{x} が連続であれば**確率密度**を表し、離散であれば**確率**を表すが、これらを統一的に扱うために、あえて小文字の $p(\cdot)$ を用いる。これに対して、離散確率変数の確率に限定されている場合には、 $P(c)$ のように大文字の $P(\cdot)$ を用い、確率密度の場合には $f(\mathbf{x})$ のように小文字の $f(\cdot)$ を用いる。

一度このような d 次元特徴空間上の点の分類問題として定式化されれば、その後は、一般的な問題としてパターン認識を扱うことができる。しかしながら、実際の応用場面においては、この特徴空間をどのように構成するかにより、パターン識別の性能に大きな影響が出る。この特徴抽出をどのようにうまく設計するかによって、認識性能の大半が決定されてしまうといっても過言ではない。しかしながら、有効な特徴抽出の設計法は、それぞれの対象問題に依存しており、普遍的に優れた方法は考えにくい。対象問題に精通した技術者が、さまざまな泥臭い方法を駆使して、有効な特徴空間を作り上げるのが普通である。

1.2 ベイズ識別規則と識別関数法

一般に、あるクラスから生成される特徴ベクトルは不確実性を有しており、パターン認識の一般的なモデルとして、これを確率分布で表現することはよく行われている。ここでは、ある確率分布に従ってデータが生成されると仮定できるとして、合理的な判別方法を与えてみよう。

1.2.1 ベイズ決定による最適分類

同じクラスに属するパターンの集合を同じ確率分布に従う実現値と見なし、認識すべき入力パターンがどのクラスの確率分布から発生したと考えるのが最も妥当かを検討すればよい。従来、このような問題は統計的仮説検定や統計的決定問題として扱われてきた。

ここで、クラス c' から発生したパターンを間違えてクラス c に属していると判断したときの損失を考える。この損失関数を $l(c, c')$ で表すと、入力パターン \mathbf{x} に対する損失の期待値は

$$L(c|\mathbf{x}) = \sum_{c' \in C} l(c, c')P(c'|\mathbf{x}) \quad (1.3)$$

となる。これを最小にする c が \mathbf{x} の属するクラスであると決定する方法は、平均損失を最小化するという意味で最適である。この方法をベイズ識別法 (Bayes discriminant method) という。

$P(c|\mathbf{x})$ は条件付き確率であり、事前確率 $P(c)$ と、クラス c のもとでの特徴ベク

トル \mathbf{x} の生起確率分布 $p(\mathbf{x}|c)$ とを用いて

$$P(c|\mathbf{x}) = \frac{p(\mathbf{x}|c)P(c)}{p(\mathbf{x})} \quad (1.4)$$

で与えられる。ただし、ここでも \mathbf{x} が離散と連続の双方の場合があることを考慮して、 \mathbf{x} の確率分布として $p(\mathbf{x}|c)$ という小文字の記号をあえて用い、離散値であるクラスの事前確率と事後確率は、大文字で $P(c)$, $P(c|\mathbf{x})$ と記述している。

ここで、損失関数として 0-1 損失

$$l(c, c') = \begin{cases} 0 & (c = c' \text{ のとき}) \\ 1 & (c \neq c' \text{ のとき}) \end{cases} \quad (1.5)$$

を仮定してみる。この損失関数は非常に取り扱いやすい性質を有しており、このとき平均損失関数は

$$L(c|\mathbf{x}) = 1 - P(c|\mathbf{x}) \quad (1.6)$$

となり、この最小化は $P(c|\mathbf{x})$ の最大化と等価である。したがって、この場合の識別を**事後確率最大のベイズ識別**という。

式 (1.4) の分母の $p(\mathbf{x})$ はクラスによらないため、事後確率最大のベイズ識別では

$$\hat{c} = \arg \max_{c \in \mathcal{C}} p(\mathbf{x}|c)P(c) \quad (1.7)$$

または

$$\hat{c} = \arg \max_{c \in \mathcal{C}} \{\log p(\mathbf{x}|c) + \log P(c)\} \quad (1.8)$$

によって \mathbf{x} の属するクラスを \hat{c} と判定すればよい (\hat{c} は c の推定値)。さらに、クラスの事前確率 $P(c)$ がすべての c で等しい一様分布の場合には

$$\hat{c} = \arg \max_{c \in \mathcal{C}} p(\mathbf{x}|c) \quad (1.9)$$

となる。

条件付き確率 $p(\mathbf{x}|c)$ は、クラス c において特徴パターン \mathbf{x} が生成される確率であり、これを計算するためには、特徴パターンの分布をクラスごとに確率モデルで表現する必要がある。実際のパターン認識では、特徴パターンがある点を中心に分布している場合が多く見られ、多次元正規分布によって特徴パターンの分布を近似で

索引

【あ】		学習定数	117	結合重み	112
赤池情報量基準	212	学習データ	13	決定木	15, 72
誤り訂正符号	172	確率的降下法	40, 115	【こ】	
アンサンブル	15	確率的コンプレキシティ	222	交互作用	74
【い】		カテゴリ	1	交差検証法	82
イェンセンの不等式	197	完全データ	198	勾配ベクトル	37
一貫性	218	ガンマ関数	27, 190	硬判定	175
一般化デルタルール	114	関連ベクトルマシン	155	誤差逆伝播学習法	114, 118
入れ子型モデル族	206, 208	【き】		子ノード	73
【う】		記憶に基づく学習	13, 60	誤分類率	75
ウエイト	12, 29, 112	擬距離	187	混合正規分布	195
【え】		木構造	72	混合正規モデル	194
枝	73	記述長最小基準	216	混合モデル	14, 194
エッジ	73	基準化	19	混合 Poly α 分布	202
エントロピー	75	逆行列補題	155	【さ】	
【お】		強一貫性	218	最急降下法	139, 148
重み係数	12	共役事前分布	191	最急降下方向	115
親ノード	73	教師付き学習	5	最近傍識別法	65
【か】		教師なし学習	5	最小 2 乗法	123
カーネル	14, 121	共分散行列	20	最尤推定量	181, 182
カーネル関数	121, 127, 141	行列式補題	157	サポートベクトル	
カーネル法	16, 121	距離	187		135, 139, 147
カーネルリッジ回帰	128	【く】		サポートベクトルマシン	14, 134, 135
カーネル SVM	144	区分的線形識別関数	110	算術平均	182
回帰木	72	クラス	1	サンプルスコア	54
階層モデル族	206	クラス間分散	24	【し】	
概念	1	クラスタ分析	5	しきい値	111
ガウスクーネル	128, 143	クラスターリング	5	識別	4
過学習	119, 211	クラス内分散	24	識別関数	8
学習	2	グラム行列	128	識別関数法	8
【け】		【く】		識別境界	79, 125, 135
経験分布	214	区分的線形識別関数	110	シグモイドカーネル	131
計量行列	69	クラス	1		

シグモイド関数 113
次元縮約 47
次元の呪い 25
事後確率最大推定量 186
事後確率最大のベイズ識別 7, 179
事後確率分布 184
指数型損失関数 91
事前確率分布 184
自然言語 203
自然言語処理 203
弱一致性 218
集団学習法 85
自由度 183
主成分得点 54
主成分分析 47
情報利得 76
情報量基準 220

【す】

推定誤差 101
スタンプ 94
ステップ関数 113
スパース 43
スペクトル分解 56

【せ】

正規化カーネル 131
生成行列 173
制約付き最適化問題 137
絶対誤差損失 185
節 点 73
漸近正規性 214
線形カーネル 131
線形サポートベクトルマシン 141
線形識別関数 11, 29
線形識別モデル 14
線形分離可能 30
線形分離不可能 31
潜在クラスモデル 14, 194
潜在トピック 201
潜在トピックモデル 203

【そ】

相関係数 21
相互情報量 76
属性番号 74
ソフトマージンサポートベクトルマシン 144, 145
損失関数 184
損失最小推定量 209

【た】

対角化 56
対数尤度関数 181
代表ベクトル 20, 63
竹内の情報量基準 215
多項式カーネル 131
多項分布 201
多値分類問題 169
多峰形 186
単峰形 186

【ち】

知識情報処理 2
中間ノード 73
中心極限定理 214
直交化学習 39

【て】

ディープラーニング 14
ディリクレ分布 202
手書き文字認識 2
テキストデータ 202
デルタルール 39
テンプレート 61
テンプレートマッチング 13, 14, 61

【と】

統計的学習 2, 13
統計的推定問題 181
統計的モデル選択 205
統計的モデル選択問題 206
特徴抽出 4, 18
特徴ベクトル 4, 5

特徴量 4

【な】

ナイーブベイズ法 14, 41
内 積 62
軟判定 175

【に】

二値分類問題 169
入出力関数 113
ニューラルネットワーク 14, 112
ニューロン 111, 112

【ね】

根ノード 73

【の】

ノード 73
ノンパラメトリックな手法 15, 180

【は】

パーセプトロン 12, 30
ハードマージンサポートベクトルマシン 145
バギング 85
パターン識別器 22
パターン認識 1
パターン分類問題 4
葉ノード 73
ハミング距離 175
パラメトリック確率モデル 15
パラメトリックな手法 180
汎化能力 120
判 別 4
判別器 4
判別分析 33

【ひ】

ヒストグラム 15
微分エントロピー 188
標本分散 182

【ふ】		ベータ分布	190	ユークリッドノルム	5
フィッシャーの線形判別法	33	ベルヌーイ 試行	189	尤度関数	181
ブースティング	90	【ま】		尤離度	81
ブートストラップサンプル	85	マージン	135	ユニット	112
不完全データ	198	マハラノビス距離	69	【よ】	
復元抽出	85	【み】		余弦尺度	62
不純度	75	未観測データ	198	【ら】	
不偏推定量	181	【め】		ラグランジュ関数	137, 145
不偏分散	183	メトリック	69	ラグランジュの未定乗数法	137, 145
フルモデル	209	メトリックラーニング	70	ランダムフォレスト	15, 85
分割指数	76	メモリベースの方法	13	【り】	
分岐選択基準	77	メモリベースの予測法	60	リスク関数	184, 213
分散共分散行列	20	【も】		リッジ回帰	121, 122
分類	4	文字認識	2	【る】	
分類器	4	モデル	13	類似度	62
分類木	72	モデル選択基準	206, 211	ルートノード	73
【へ】		モデルベースの方法	13	ルベグ積分	184
ベイズ最適な分類	178	モデルベースの予測法	60	【ろ】	
ベイズ識別	6	【ゆ】		ローカルミニマ	40
——の誤り率	65	ユークリッド距離	61, 69		
ベイズの公式	184				
ベイズモデル平均化法	15, 223				
ベイズリスク	184				

【A】		【E】		KL (Kullback-Leibler) 情報量	187, 213
AdaBoost	15, 85, 90	ECOC 法	172	k 最近傍識別法	14, 16, 64
AIC	212	EM アルゴリズム	194, 198, 199	k -NN 法	14, 16, 64
【B】		Exhaustive Code	172	【L】	
BIC	216	【G】		LDA	203
BMA	223	Gini インデックス	76	【M】	
BP アルゴリズム	118	Gini 関数	75	MDL 基準	216, 217
BP 学習アルゴリズム	114	Gini 係数	76	【N】	
BT (Bradley-Terry) モデル	171	【K】		NN 法	65
【C】		KKT 条件	138, 147	【O】	
CART	72, 74	KL (Karhunen-Loève) 展開	47	one-vs-one 法	171
CART 学習アルゴリズム	77				

one-vs-the rest 法	170				
OOB 推定確率	101				
OOB データ	101				
【P】					
PLSI	203				
【R】					
RM (Reed-Muller) 符号	173				
RVM	155				
		【S】			
		Shannon の補助定理	188		
		Sherman-Morrison- Woodbury の公式	155		
		SMO	134, 150		
		SVM	134, 135		
		Sylvester の行列式定理	156		
		【T】			
		Tempered EM アルゴリズム	203		
				【U】	
				TIC	215
				Unigram Mixture モデル	202
				【W】	
				Widrow-Hoff の学習規則	39
				Woodbury の公式	155
				【数字】	
				0-1 損失	185
				2 乗誤差損失	185

— 著者略歴 —

後藤 正幸 (ごとう まさゆき)

1992年 武蔵工業大学 (現東京都市大学)
工学部経営工学科卒業
1994年 武蔵工業大学大学院工学研究科修士
課程修了 (経営工学専攻)
1997年 早稲田大学助手
2000年 博士 (工学) (早稲田大学)
2000年 東京大学助手
2002年 武蔵工業大学助教授
2008年 早稲田大学准教授
2011年 早稲田大学教授
現在に至る

小林 学 (こばやし まなぶ)

1994年 早稲田大学理工学部工業経営学科卒業
1996年 早稲田大学大学院理工学研究科修士
課程修了 (機械工学専攻)
1998年 早稲田大学助手
2000年 博士 (工学) (早稲田大学)
2001年 早稲田大学理工学総合研究センター
研究員
2002年 湘南工科大学講師
2008年 湘南工科大学准教授
2014年 湘南工科大学教授
現在に至る

入門 パターン認識と機械学習

Introduction to Pattern Analysis and Machine Learning

© Masayuki Goto, Manabu Kobayashi 2014

2014年4月30日 初版第1刷発行



検印省略

著者 後藤 正 幸
小林 学
発行者 株式会社 コロナ社
代表者 牛来真也
印刷所 三美印刷株式会社

112-0011 東京都文京区千石 4-46-10

発行所 株式会社 コロナ社

CORONA PUBLISHING CO., LTD.

Tokyo Japan

振替 00140-8-14844・電話 (03) 3941-3131 (代)

ホームページ <http://www.coronasha.co.jp>

ISBN 978-4-339-02479-1 (新宅) (製本: 愛千製本所) G

Printed in Japan



本書のコピー、スキャン、デジタル化等の無断複製・転載は著作権法上での例外を除き禁じられています。購入者以外の第三者による本書の電子データ化及び電子書籍化は、いかなる場合も認めておりません。

落丁・乱丁本はお取替えいたします