

Natural Language Processing Series

自然言語処理シリーズ

1

言語処理のための 機械学習入門

工学博士 奥村 学 監修
博士(工学) 高村 大也 著

コロナ社

刊行のことば

人間の思考、コミュニケーションにおいて不可欠なものである言語を計算機上で扱う自然言語処理という研究分野は、すでに半世紀の歴史を経るに至り、技術的にはかなり成熟するとともに、分野が細かく細分化され、また処理対象となるものも、新聞以外に論文、特許、WWW上のテキストなど多岐にわたり、さらに応用システムもさまざまなものが生まれつつある。そして自然言語処理は、現在ではWWWの普及とともに、ネットワーク社会の基盤を支える重要な情報技術の一つになりつつあるといえる。

これまでの自然言語処理に関する専門書は、自然言語処理全般を広く、浅く扱う教科書（入門書）以外には、情報検索、自動要約などを扱う、わずかな書籍が出版されているだけという状況であった。もちろん入門書の中には優良なものでは多くの読者を集めているものもあるが、一方で、拙著「テキスト自動要約」などは、読者層がより小さいと考えられるにもかかわらず、多くの読者を得ている。この現状を鑑^{かんが}みるに、読者は、「実際にいま役に立つ本」、「いまの話題に即した本」を求めているのではないかと推測される。そこで、これまでの自然言語処理に関する専門書では扱われておらず、なおかつ、「いま重要と考えられ、今後もその重要さが変わらない」と考えられるテーマを扱った書籍の出版を企画することになった。

このような背景のもと生まれた「自然言語処理」シリーズの構成を以下に示す。

1. 自然言語処理で利用される統計的手法、機械学習手法を広く扱う。

「言語処理のための機械学習入門」

近年の自然言語処理は、コーパスに基づき、統計的手法あるいは機械学習手法を用いて、規則なり知識なりを自動獲得し、それを用いた処理を行うという手法を採用することが一般的になってきている。現状、多く

ii 刊 行 の こ と ば

の研究者は、他の先端的な研究者の論文などを参考に、それらの統計的手法、機械学習手法に関する知識を得ており、体系的な知識を得る手掛かりに欠けている。海外でも、Foundations of Statistical NLP (MIT Press 刊) という教科書が 1999 年に出版され、多くの読者を得て以来、同様の本は出版されていない。そこで、そのような統計的、機械学習手法に関する体系的知識を与える専門書が必要である。

2. 情報検索，自動要約などと並ぶ，自然言語処理の応用を扱う。
 - a. 「質問応答システム」
 - b. 「情報抽出」
 - c. 「機械翻訳」

自然言語処理分野も歴史を重ね、技術もある程度成熟し、実際に使えるシステム、技術として世の中に少しずつ流通するようになってきているものも出てきている。そのようなシステム、技術として、検索エンジン、要約システムなどがあり、それらに関する書籍も出版されるようになってきている。これらと同様に、近年実用化され、また注目を集めている技術として、質問応答、情報抽出があり、これらの技術に関する書籍の必要性を感じている。

同様に、古くから研究され、すでに実用化も進んでいる技術として、機械翻訳がある。機械翻訳は、日本政府の主導する「イノベーション 25」の技術の一つとしても位置づけられている。しかし、古くから研究され、すでに実用化も進んでいる機械翻訳に関する体系的な書籍がこれまでなかったことも驚くべきことである。

3. 処理対象が新しい自然言語処理を扱う。
 - a. 「特許情報処理：言語処理的アプローチ」
 - b. 「Web 言語処理」

自然言語処理の対象とするテキストは、近年多様化し始めており、その中でも、注目を集めているコンテンツに、特許（知的財産）、WWW 上のテキストが挙げられる。これらを対象とした自然言語処理は、その

処理結果により有用な情報が得られる可能性が高いことから、研究者が加速度的に増加し始めている。しかし、これらのテキストを対象とした自然言語処理は、これまでの自然言語処理と異なる点が多く、これまでの書籍では扱われていない内容が多い。

以上、本シリーズの構成を述べてきたが、この構成は現在まだ仮のものであることを最後に付記しておきたい。2. のカテゴリーには「対話システム」が含まれてしかるべきであるし、3. のカテゴリーになぜバイオインフォマティクスとも関連する「バイオ言語処理」のようなものがないのかと思われる読者もおられるにちがいない。今後これらの候補も含め、新たな書籍が本シリーズに加わり、本シリーズがさらに充実したものとなることを祈っている。

本シリーズは、その分野の第一人者の方々に各書籍の執筆をご快諾願えたことで、成功への最初の一步を踏み出せたのではないかと考えている。シリーズの書籍が、読者がその分野での研究を始めるうえで役に立ち、また実際のシステム開発のうえで参考になるとしたら、この企画を始めた者として望外の幸せである。最後に、このような画期的な企画にご賛同下さり、実現に向けた労をとって下さったコロナ社の各氏に感謝しておきたい。

2009年6月

監修者 奥村 学

ま え が き

言語処理に機械学習が積極的に取り入れられるようになって久しい。今日、学会などで提案される新しい手法の多くでは、その一部分あるいは大部分において機械学習を用いており、非常に高性能な処理を実現している。またその威力に魅せられ、機械学習を用いて言語処理のアプリケーションを構築している技術者の方々も多くいるだろう。もちろん、機械学習は決して言語処理のすべてではない。言語学の立場から言語処理に取り組んでおられる研究者の方々も多い。しかし、そのようなの方々にとっても機械学習は無視できない存在になっていることは事実である。

ところが、機械学習を理解することは容易ではない。用語の意味もぼんやりとしかわからなく、論文や解説書が理解できず、それでは基礎的なことから勉強しようと思ってもどこから勉強したらいいかわからないし、機械学習以外のこともたくさん勉強しなくてはいけないので時間もない、というような状況はよくあるのではないだろうか。

このような状況を踏まえ、本書は、機械学習を用いた言語処理技術を理解するための基礎的な知識や考え方を伝えることを目的としている。特に、これから言語処理を志す大学生や大学院生を含む初学者の方々、これまで機械学習をそれほど用いてこなかった言語処理研究者の方々、もう一度機械学習を見直したい言語処理技術者の方々などの助けになるように書いたつもりである。もちろん、すべてを一冊に詰め込むことは不可能であるので、必要な知識だけを拾って、できる限り一直線に、実際に使われている機械学習手法へたどり着くように説明していく。これにより、数学的な概念と機械学習手法がどのようにつながっているかが、わかるようになるだろう。

今後、言語処理分野から多くの優れた論文や技術が生まれること、またそれらが実用化され社会に役立つことを期待している。そして、本書がその一助を担うことができれば幸いである。

本書を著すにあたり、多くの方々にお世話になった。東京工業大学教授の奥村

学先生には、本書のご提案と監修をいただいたのみならず、内容や構成について多くの助言をいただいた。ヤフー株式会社の要素技術開発部の若手の方々や甲南大学の永田亮先生には、改善点に関するたくさんの指摘をいただいた。東京工業大学の森田一君には、章末問題の解答の誤りを発見していただいた。また、コロナ社様には、本書の刊行のためにたいへんご尽力いただいた。最後に、家族や友人にはつねに励まされそして支えられた。ここに感謝の意を表したい。

2010年6月

高村大也

目 次

1. 必要な数学的知識

1.1	準備と本書における約束事	1
1.2	最適化問題	4
1.2.1	凸集合と凸関数	6
1.2.2	凸計画問題	12
1.2.3	等式制約付凸計画問題	14
1.2.4	不等式制約付凸計画問題	18
1.3	確率	20
1.3.1	期待値, 平均, 分散	23
1.3.2	結合確率と条件付き確率	26
1.3.3	独立性	29
1.3.4	代表的な離散確率分布	31
1.4	連続確率変数	37
1.4.1	平均, 分散	38
1.4.2	連続確率分布の例	38
1.5	パラメータ推定法	41
1.5.1	i.i.d. と尤度	42
1.5.2	最尤推定	42
1.5.3	最大事後確率推定	46
1.6	情報理論	49
1.6.1	エントロピー	49
1.6.2	カルバック・ライブラー・ダイバージェンス	51

1.6.3	ジェンセン・シャノン・ダイバージェンス	54
1.6.4	自己相互情報量	56
1.6.5	相互情報量	57
1.7	この章のまとめ	58
章末問題		59

2. 文書および単語の数学的表現

2.1	タイプ, トークン	61
2.2	n グラム	62
2.2.1	単語 n グラム	62
2.2.2	文字 n グラム	63
2.3	文書, 文のベクトル表現	64
2.3.1	文書のベクトル表現	64
2.3.2	文のベクトル表現	66
2.4	文書に対する前処理とデータスパースネス問題	67
2.4.1	文書に対する前処理	67
2.4.2	日本語の前処理	70
2.4.3	データスパースネス問題	71
2.5	単語のベクトル表現	71
2.5.1	単語トークンの文脈ベクトル表現	72
2.5.2	単語タイプの文脈ベクトル表現	73
2.6	文書や単語の確率分布による表現	74
2.7	この章のまとめ	75
章末問題		76

3. クラスタリング

3.1 準 備	77
3.2 凝集型クラスタリング	78
3.3 k -平均法	82
3.4 混合正規分布によるクラスタリング	84
3.5 EM アルゴリズム	87
3.6 クラスタリングにおける問題点や注意点	94
3.7 この章のまとめ	95
章 末 問 題	97

4. 分 類

4.1 準 備	99
4.2 ナイーブベイズ分類器	101
4.2.1 多変数ベルヌーイモデル	102
4.2.2 多項モデル	110
4.3 サポートベクトルマシン	117
4.3.1 マージン最大化	118
4.3.2 厳密制約下の SVM モデル	120
4.3.3 緩和制約下の SVM モデル	124
4.3.4 関数距離	125
4.3.5 多値分類器への拡張	126
4.4 カーネル法	127
4.5 対数線形モデル	132
4.5.1 素性表現の拡張と対数線形モデルの導入	132
4.5.2 対数線形モデルの学習	134

4.6 素性選択	138
4.6.1 自己相互情報量	138
4.6.2 情報利得	141
4.7 この章のまとめ	143
章末問題	145

5. 系列ラベリング

5.1 準備	147
5.2 隠れマルコフモデル	148
5.2.1 HMMの導入	149
5.2.2 パラメータ推定	149
5.2.3 HMMの推論	150
5.3 通常のカテゴリの逐次適用	151
5.4 条件付確率場	153
5.4.1 条件付確率場の導入	153
5.4.2 条件付確率場の学習	154
5.5 チャンキングへの適用の仕方	159
5.6 この章のまとめ	160
章末問題	161

6. 実験の仕方など

6.1 プログラムとデータの入手	162
6.2 分類問題の実験の仕方	163
6.2.1 データの分け方と交差検定	164
6.2.2 多クラスと複数ラベル	165
6.3 評価指標	166

6.3.1	分類正解率	166
6.3.2	精度と再現率	166
6.3.3	精度と再現率の統合	168
6.3.4	多クラスデータを用いる場合の実験設定	170
6.3.5	評価指標の平均	171
6.3.6	チャンキングの評価指標	175
6.4	検定	175
6.5	この章のまとめ	177
	章末問題	178
	付 録	180
A.1	初歩的事項	180
A.2	logsumexp	183
A.3	カルーシュ・クーン・タッカー (KKT) 条件	184
A.4	ウェブから入手可能なデータセット	185
	引用・参考文献	186
	章末問題解答	190
	索 引	206

1

必要な数学的知識

この章で紹介したい事柄は、最適化問題、確率論、情報理論の三つである[†]。最初に 1.1 節で、簡単な準備と本書における記法などの約束事の確認を行う。残りの節は数学的な基礎知識に関するものである。1.2 節で説明する最適化問題では、特に制約付き最適化問題の解き方（ラグランジュ乗数法）の理解を目指す。ラグランジュ乗数法は、言語処理において確率モデルや機械学習を用いるうえで必要不可欠なテクニックである。1.3 節から 1.5 節では確率について説明する。次章以降で説明する多くの手法が確率に基づいており、決して避けて通れない内容である。特に、言語処理で頻繁に使用する離散確率について重点的に説明する。1.6 節は情報理論に関する節である。情報理論における、エントロピー、カルバック・ライブラー・ダイバージェンス、ジェンセン・シャノン・ダイバージェンス、自己相互情報量など、言語処理で用いられる指標とその意味を概説することにしよう。

1.1 準備と本書における約束事

最初に、言語処理におけるおもなタスクを概観しておこう。まず、単語分割 (word segmentation) は、与えられた文を単語に分割するタスクである。英語などの言語では単語間に空白があるので、基本的には単語分割は不要である。つぎに、品詞タグ付け (part-of-speech tagging) というタスクでは、文中の単語の品詞を推定する。字面が同じでも品詞が異なることはあるので、高精度な品詞タグ付けはみかけほど簡単ではない。構文解析 (syntactic parsing) では、

[†] この三つの事柄を説明するためのさらに初歩的な事柄は、付録 A.1 に記しておいたもので、必要に応じて参照していただきたい。基礎に不安が残る方は、そちらを先に読んでいただいてもよい。

2 1. 必要な数学的知識

文の構文的な構造を推定する。以上は、文の構造をさまざまなレベルで明らかにするタスクであったが、これ以外にも多くのタスクが存在する。文書分類 (text classification) では、与えられた文書に対して、いくつかのクラスのうちの適切なものを決定する。クラスとしては、内容に着目したものもあれば (スポーツ, 政治, 経済など), 筆者ごとにクラスを考えるものもあり (芥川龍之介クラス, 夏目漱石クラス, 太宰治クラス), じつにさまざまな文書分類タスクが考えられる。さらには, 照応解析, 質問応答, 機械翻訳などタスクは枚挙に暇がない。それぞれのタスクでの対象の処理単位を事例 (instance) とよぶ。例えば, 文書分類では各文書が事例である。

機械学習とはデータから分類規則などを学ぶ枠組みである。つまりデータがなければはじまらない。言語処理では多くの場合, 言語のさまざまな形の用例の集まりである, コーパス (corpus) とよばれる言語データを用いる。例えば辞書などはもちろん言語データであるが用例ではないので, 一般的にはコーパスではないとされる。機械学習を用いた言語処理では, コーパスからいろいろな数値を取り出し, 数学の世界に持ち込んで, そこでなんらかの処理を行うことになる。

数学の世界に持ち込むために, 本書では数式を用いた説明を避けることができない。しかし, 直感的理解を大事にしたいという考えから, 数式の記述においても, 目に飛び込んでくる記号や文字をできる限り減らした。このため, 厳密とはいえない説明がときどき出てくるだろう。例えば, 定義域を厳密に記述しなかったり, きわめて例外的と思われる状況については無視したりしている。また一方, 平易な数式変形を妙に詳しく説明したり, 自明と思われる例を多用したりもしている。これも, 理解の助けになると信じてのことである。

さて, 本書での約束事をここで少し述べておく。まず, 高校で習うレベルの数学の基礎知識 (簡単な微分演算, 確率) は前提とする。もちろん必要と思われる部分については, 本書の中で簡単に復習する。また本書でさまざまな関数が登場するが, それらの微分可能性などはあらかじめ仮定しておくことにしよう。

つぎに, 数式に用いる記号などの使い方について述べる。不要だと思われた

らとばしてほしい。本書での文字の使い方としては、 x, y, z などを優先的に変数として用いる。 x_1, x_2, x_3, x_4 などで変数を表すこともある。例えば、変数がたくさんあるときや、ベクトル \mathbf{x} の要素であるような場合などに、こちらを好んで使う。同じように変数と名のつくものに確率変数があるが、これは X など大文字で表すことにし、確率変数の値は x など小文字で表すことにする。つまり、小文字のアルファベットは、状況により変数を表したり決まった値を表したりする。紛らわしいが、正確に読み解く訓練をしてほしい。また、データとして与えられた事例は、 $x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}$ などのように上付の添字とともに表すことにしておく。これらはサンプルであるので、例えば、 $x^{(1)}$ と $x^{(2)}$ が同じ値ということもありうる。このような使い分けは、論文でもなされることがあるが、別の記法が用いられることも多い。何が何を表すかを正確に把握することは、論文を正しく読むコツでもある。

また、 $n(w, d)$ およびより簡単に $n_{w,d}$ は文書 d における単語 w の出現回数、 $n(w, s)$ および $n_{w,s}$ は文 s における単語 w の出現回数を表すものとする。 $n(w, c)$ および $n_{w,c}$ はクラス c に属する文書群における単語 w の出現回数を表す。似て非なるものとして、 $N(w, c)$ および $N_{w,c}$ があり、これはクラス c に属する文書のうち w が出現するような文書の数を表す。 $N(c)$ および N_c はクラス c に属する文書数を表す。ただし、 n や N における d, s などは、文脈から明らかである場合は省略することもある。

また、 $\delta(w, d)$ およびより簡単に $\delta_{w,d}$ は、文書 d において単語 w が出現したら 1、そうでなければ 0 となる。 $\delta(w, s)$ や $\delta_{w,s}$ なども、文に対して同様の意味を持つものとする。 d や s を省略することがあるのは、 δ についても同じである。

このような記法を用いて説明を行っていく。当然ながらこういった記法は文献により異なるが、本書で採用する記法は多数の文献で採用されているものである。

次節からは数学的な基礎知識について説明していく。

1.2 最適化問題

最適化問題 (optimization problem) とは、ある制約のもとで関数を最適化 (最大化あるいは最小化) する変数値とそのときの関数値を求める問題である。それぞれ、**最大化問題** (maximization problem)、**最小化問題** (minimization problem) とよばれる。例を挙げてみよう。ただし、例題中の $\max.$ は maximize の略で、これが最大化問題であることを示しており、最小化問題ならば $\min.$ になる。

例題 1.1 つぎの最大化問題を解け (ただし、 a は定数である) :

$$\begin{aligned} \max. \quad & -x_1x_2 \\ \text{s.t.} \quad & x_1 - x_2 - a = 0. \end{aligned}$$

【解答】 これは、つぎのように非常に簡単に解ける。制約から $x_2 = x_1 - a$ として、目的関数に代入する :

$$\begin{aligned} -x_1x_2 &= -x_1(x_1 - a) \\ &= -x_1^2 + ax_1. \end{aligned}$$

x_1 で偏微分して 0 とおけば、 $x_1 = a/2$ 、 $x_2 = -a/2$ が得られる。 \diamond

この例題では、 $-x_1x_2$ という関数を最大化した。このように最適化したい関数をこの最適化問題の**目的関数** (objective function) という。また、上の例題での $(x_1, x_2) = (a/2, -a/2)$ のように最適値を与える変数値を**最適解** (optimal solution) という。

一般には最適化問題はつぎのように書かれる :

$$\max. \quad f(x) \tag{1.1}$$

$$\text{s.t.} \quad g(x) \geq 0 \tag{1.2}$$

$$h(x) = 0. \tag{1.3}$$

ここで、 $f(x)$ が目的関数である。また、 $g(x) \geq 0$ 、 $h(x) = 0$ が制約である。特に、 $g(x) \geq 0$ を不等式制約 (inequality constraint)、 $h(x) = 0$ を等式制約 (equality constraint) とよぶ。「s.t.」以下に制約を記述するのが一般的であり、省略せずに subject to が用いられることもある。制約を満たす解のことを実行可能解 (feasible solution) という。上の例題では、 $x_1 - x_2 - a = 0$ を満たすような解が実行可能解である。また、実行可能解の集合を実行可能領域 (feasible region) という。ちなみに、 $\max.f(x)$ と $\min.-f(x)$ は等価であるので、両者の間に本質的な違いはない。

さて、上の例題では $x_1 = a/2$ 、 $x_2 = -a/2$ と答が得られた。このような、「 $x_1 =$ 」などのような解の表し方、もう少し詳しくいうと、加減乗除や初等関数の合成関数による解の表し方を、閉形式 (closed-form) という^{†1}。すべての問題について閉形式の解が得られるわけではない。閉形式の解が得られる問題を、解析的に解ける (analytically solvable) 問題という^{†2}。

さて、本節のこれ以降の部分では、最適化問題の中の凸計画問題 (convex programming problem) とよばれる問題を取り扱う。具体的にどのような問題であるかはこれから説明していくが、大雑把にいうと、凸計画問題では、目的関数の値が改善する方向に進んでいけば解にたどりつく。それゆえ、凸計画問題は比較的解きやすい問題であり、その解法はある程度確立している。つまり、解きたい問題を凸計画問題で表すことに成功すれば、問題解決に向けて大躍進したことになる。凸計画問題で表すことができるような問題だけでも非常に多くのことができるし、凸計画問題でない問題を解くアルゴリズムでも、実際には内部で凸計画問題を繰り返し解いていたりする。

以下で、凸計画問題をもう少し詳しく説明する。まず、凸計画問題を定義するために、1.2.1 項で凸集合および凸関数という二つの概念を導入する。つぎに、1.2.2 項で凸計画問題を定義し、制約がない場合の数値解法について簡単に述べ

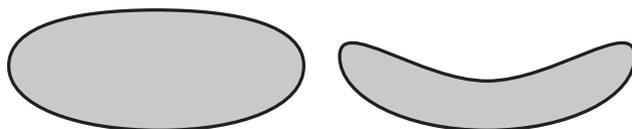
^{†1} 微分形式でいう閉形式とは異なる。

^{†2} ただし、初等関数としてどれだけの関数を含めるかは、明確に定められていないので、閉形式の定義は曖昧である。しかし、この曖昧さは実用上はほとんど問題にならない。

る。さらに 1.2.3 項で等式制約付凸計画問題, 1.2.4 項で不等式制約付凸計画問題について説明する。

1.2.1 凸集合と凸関数

凸集合とは, 直感的にいえば「へこみのない集合」である。図 1.1 (a) のような集合は凸集合であるが, 図 (b) のような集合は凸集合ではない。



(a) 凸集合

(b) 凸ではない集合

図 1.1 凸集合と凸ではない集合

このことを厳密に記述してみよう。集合 $A \subseteq R^d$ が凸集合 (convex set) であるとは^{†1}, 任意の $\mathbf{x}^{(1)} \in A$ と $\mathbf{x}^{(2)} \in A$ があつたとき, 任意の $t \in [0, 1]$ に対して $t\mathbf{x}^{(1)} + (1-t)\mathbf{x}^{(2)} \in A$ が成り立つことである。 $t\mathbf{x}^{(1)} + (1-t)\mathbf{x}^{(2)}$ は, $t \in [0, 1]$ のとき $\mathbf{x}^{(1)}$ と $\mathbf{x}^{(2)}$ を結ぶ線分を表しているので (付録 A.1.2 の〔1〕), この条件は, 集合内の任意の点を結ぶ線分が集合 A 自身からはみ出ないことを表している。つまり, へこみが無いということである^{†2}。

例題 1.2 ある与えられた平面上にあるという制約を満たすベクトルの集合 $A = \{\mathbf{x} | \mathbf{m} \cdot \mathbf{x} + b = 0, \mathbf{x} \in R^d\}$ は凸集合であることを示せ。

【解答】 $\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in A$ とする。つまり, $\mathbf{m} \cdot \mathbf{x}^{(1)} + b = 0$ かつ $\mathbf{m} \cdot \mathbf{x}^{(2)} + b = 0$ である。任意の実数 $t \in [0, 1]$ に対して, $t\mathbf{x}^{(1)} + (1-t)\mathbf{x}^{(2)}$ もその平面上にあることを示せばよい。実際に平面の方程式に代入してみると,

$$\begin{aligned} & \mathbf{m} \cdot (t\mathbf{x}^{(1)} + (1-t)\mathbf{x}^{(2)}) + b \\ &= t\mathbf{m} \cdot \mathbf{x}^{(1)} + (1-t)\mathbf{m} \cdot \mathbf{x}^{(2)} + b \end{aligned}$$

^{†1} 要素が実数であるような d 次元のベクトルの集合を R^d で表す。

^{†2} 凸集合は R^d の部分集合である必要はないが, 簡単のためにここでは R^d の部分集合であるとしておく。

索引

<p>【あ】</p> <p>i.i.d. 42</p> <p>IOB2 タグ値 159</p> <p>180</p> <p>RF カーネル 130</p> <p>鞍点 18</p> <p>【い】</p> <p>EM アルゴリズム 87</p> <p>【う】</p> <p>ウィルコクソンの符号付順位和検定 177</p> <p>【え】</p> <p>HMM 148</p> <p>Expectation-Maximization アルゴリズム 87</p> <p>n グラム 62</p> <p>F 値 168</p> <p>エントロピー 49</p> <p>【お】</p> <p>凹関数 7</p> <p>【か】</p> <p>解析的に解ける 5</p> <p>開発データ 164</p> <p>ガウス分布 38</p> <p>学習 78</p> <p>学習データ 78</p> <p>学習率 13</p> <p>確率関数 22</p> <p>確率質量関数 22</p>	<p>確率的勾配法 137</p> <p>確率的潜在意味解析 93</p> <p>確率的潜在意味索引付け 93</p> <p>確率伝搬法 161</p> <p>確率分布 22</p> <p>確率変数 21</p> <p>確率密度関数 37</p> <p>隠れ変数 90</p> <p>隠れマルコフモデル 148</p> <p>カテゴリ 100</p> <p>カーネル関数 128</p> <p>カーネル法 128</p> <p>カルーシュ・クーン・タッカー条件 184</p> <p>カルバック・ライブラー・ダイバージェンス 51</p> <p>関数 180</p> <p>関数距離 126</p> <p>完全データ 90</p> <p>観測変数 90</p> <p>【き】</p> <p>木構造カーネル 129</p> <p>規則ベース手法 100</p> <p>期待値 23</p> <p>帰無仮説 176</p> <p>Q 関数 88</p> <p>教師付き学習 101</p> <p>教師無し学習 101</p> <p>凝集型クラスタリング 78</p> <p>【く】</p> <p>クラス 100</p> <p>クラスタ 78</p> <p>クラスタリング 78</p>	<p>クラスラベル 100</p> <p>訓練 78</p> <p>訓練事例 78</p> <p>訓練データ 78</p> <p>【け】</p> <p>KL ダイバージェンス 52</p> <p>形態素解析 70</p> <p>k-平均法 82</p> <p>系列 147</p> <p>系列ラベリング 147</p> <p>結合確率 27</p> <p>言語モデル 76</p> <p>【こ】</p> <p>交差検定 164</p> <p>構文解析 1</p> <p>語義の曖昧性解消 70</p> <p>異なり語数 62</p> <p>コーパス 2</p> <p>混合正規分布 85</p> <p>【さ】</p> <p>最急降下法 13</p> <p>最急勾配法 13</p> <p>最急上昇法 13</p> <p>再現率 167</p> <p>再現率-精度曲線 167</p> <p>再現率/精度 break-even ポイント 169</p> <p>最小化問題 4</p> <p>最大エントロピーモデル 132</p> <p>最大化問題 4</p> <p>最大事後確率推定 46</p> <p>最適解 4</p>
---	--	--

最適化問題 4
 最尤推定 43
 サポートベクトルマシン 117
 算術平均 25

【し】

CRF 153
 JS 情報量 54
 JS ダイバージェンス 54
 ジェンセン・シャノン情報量 54
 ジェンセン・シャノン・ダイバージェンス 54
 次元 180
 事後確率 85
 事後確率分布 46
 自己相互情報量 56
 事象 21
 事象空間 21
 事前確率分布 46
 実行可能解 5
 実行可能領域 5
 11 点平均精度 169
 重心ベクトル 81
 従属 30
 周辺確率 29
 樹形図 79
 主問題 19
 準ニュートン法 137
 条件付きエントロピー 51
 条件付き確率 27
 条件付き確率場 153
 条件付き確率分布 27
 条件付き独立 30
 情報利得 141
 事例 2

【す】

数値解法 12
 スカラー 180
 スカラー値関数 180
 ステミング 68
 ストップワード 68

スペクトラルクラスタリング 96
 スムージング 110

【せ】

正解率 166
 正規分布 38
 正クラス 118
 正則化 134
 精度 167
 正例 118

【そ】

疎 71
 相互情報量 57
 双対問題 19
 素性 64
 素性関数 132
 素性選択 138
 素性値 64

【た】

対数線形モデル 132
 対数尤度 42
 タイプ 62
 多クラスデータセット 165
 多項式カーネル 129
 多項分布 35
 多項モデル 110
 多値分類問題 165
 多変数ベルヌーイ分布 32
 多変数ベルヌーイモデル 102
 ダミーの単語 63
 単一ラベルデータセット 165
 単語 n グラム 63
 単語タイプ 62
 単語トークン 62
 単語分割 1

【ち】

逐次最小最適化 123
 チャンキング 159

【て】

t -検定 177
 デリクレ分布 39
 データスパースネス問題 71

【と】

動径基底関数カーネル 130
 統計的検定 175
 統計的に有意 176
 同時確率 27
 等式制約 5
 トークン 62
 特徴 64
 特徴量 64
 独立 30
 独立に同一の確率分布に従う 42
 凸関数 7
 凸関数であるための 1 次条件 10
 凸関数であるための 2 次条件 10
 凸計画問題 5
 凸集合 6
 トライグラム 62

【な】

内積 180
 ナイーブベイズ分類器 101

【に】

二項分布 33
 2 次計画問題 122
 二値クラスデータセット 165
 二値分類問題 165
 二値ベクトル 65
 ニュートン法 13

【の】

述べて語数 62

【は】	文脈ベクトル	72	マージン	119	
バイグラム	62	文脈窓	72	マージン最大化	119
バウム-ウェルチアルゴリズム	160	文脈窓幅	72	MAP 推定	46
bag-of-ngrams	66	分離平面	119	【み】	
bag-of-words	65	分類	100	見出し語化	68
半教師付き学習	144	分類器	100	【も】	
半正定値	11	分類規則	100	目的関数	4
半負定値	11	分類正解率	166	文字 n グラム	63
		【へ】		文字列カーネル	129
【ひ】		ペアワイズ法	127	【ゆ】	
非観測変数	90	平均	23	有意	176
引数	180	平均ベクトル	25	有意水準	176
ヴィタピアルゴリズム	150	閉形式	5	尤度	42
p 値	176	ベイズ推定	58	ユニグラム	62
評価事例	164	ベイズの定理	28	【ら】	
評価データ	164	ベクトル	180	ラグランジュ関数	14
標本空間	31	ベクトル値関数	180	ラグランジュ乗数	14
標本分散	26	ヘッセ行列	11	ラグランジュの乗数法	15
標本平均	25	ベルヌーイ分布	31	ラベル	100
品詞タグ付け	1	偏微分	180	ラベル付きデータ	100
頻度ベクトル	65	【ほ】		ラベル無しデータ	101
		ポアソン分布	36	【り】	
【ふ】		方向ベクトル	182	離散確率変数	22
不完全データ	90	法線ベクトル	183	【れ】	
複数ラベルデータセット	165	ポーターのステマー	68	連続確率変数	22, 37
負クラス	118	ボトムアップクラスタリング	79	連続変数	37
符号検定	177	【ま】		【わ】	
不等式制約	5	マイクロ平均	172	one-versus-rest 法	126
負例	118	前向き・後ろ向きアルゴリズム	157		
プロダクトモデル	98	マクロ平均	172		
分割表	167				
分散	24				
文書分類	2				

【A】	arithmetic mean	25	Bayesian inference	58	
accuracy	166	attribute	64	Bayes' theorem	28
agglomerative clustering	78	attribute value	64	belief propagation	161
analytically solvable	5	【B】		Bernoulli distribution	31
argument	180	Baum-Welch algorithm	160	bigram	62
				binary classification	

problem	165	CRF	153		
binary vector	65	cross-validation	164		
binary-class dataset	165			[G]	
binomial distribution	33	[D]		Gaussian distribution	38
bottom-up clustering	79	data sparseness problem	71	Gaussian mixture	85
		dendrogram	79	gradient ascent method	13
[C]		dependent	30	gradient descent method	13
categorization	100	development data	164	gradient method	13
category	100	dimension	180		
centroid	81	direction vector	182	[H]	
character n -gram	63	Dirichlet distribution	39	Hessian	11
chunking	159	discrete random variable	22	hidden Markov model	148
class	100	dual problem	19	HMM	148
classification	100	dummy word	63		
classification accuracy	166			[I]	
classification rule	100	[E]		i.i.d.	42
classifier	100	eleven point average		incomplete data	90
class label	100	precision	169	independent	30
closed-form	5	EM algorithm	87	independently, identically	
cluster	78	entropy	49	distributed	42
clustering	78	equality constraint	5	inequality constraint	5
complete data	90	event	21	information gain	141
concave	7	event space	21	inner product	180
conditionally independent		Expectation-Maximization		instance	2
	30	algorithm	87	IOB2 tag	159
conditional entropy	51	expected value	23		
conditional probability	27			[J]	
conditional probability		[F]		Jensen-Shannon	
distribution	27	feasible region	5	divergence	54
conditional random		feasible solution	5	joint probability	27
fields	153	feature	64	JS divergence	54
context vector	72	feature function	132		
context window	72	feature selection	138	[K]	
context window size	73	feature value	64	Karush-Kuhn-Tucker	
contingency table	167	first-order convexity		condition	184
continuous random		condition	10	kernel function	128
variable	37	forward-backward		kernel method	128
continuous variable	37	algorithm	157	KL divergence	52
convex function	7	frequency vector	65	Kullback-Leibler	
convex programming		function	180	divergence	52
problem	5	functional distance	126	k -means	82
convex set	6	F-measure	168		
corpus	2			[L]	
				label	100

sequential labeling	147	SVM	117	unobserved variable	90
sequential minimal optimization	123	syntactic parsing	1	unsupervised learning	101
significance level	176				
significant	176	[T]		[V]	
sign test	177	test data	164	value	180
single-label dataset	165	test instance	164	variance	24
SMO	123	text classification	2	vector	180
smoothing	110	the method of Lagrange multipliers	15	vector function	180
sparse	71	token	62	Viterbi algorithm	150
spectral clustering	96	training	78		
statistically significant	176	training data	78	[W]	
statistical test	175	training instance	78	Wilcoxon's signed rank sum test	177
stemming	68	tree kernel	129	word segmentation	1
stochastic gradient method	137	trigram	62	word sense disambiguation	70
stopword	68	type	62		
string kernel	129	<i>t</i> -test	177	word token	62
supervised learning	101			word type	62
Support Vector Machine	117	[U]		word <i>n</i> -gram	63
		unigram	62		
		unlabeled data	101		

監修者・著者略歴

奥村 学 (おくむら まなぶ)	高村 大也 (たかむら ひろや)
1984年 東京工業大学工学部情報工学科卒業	1997年 東京大学工学部計数工学科卒業
1989年 東京工業大学大学院博士課程修了 (情報工学専攻), 工学博士	2000年 東京大学大学院工学系研究科修士課程修了 (計数工学専攻)
1989年 東京工業大学助手	2003年 奈良先端科学技術大学院大学情報科学研究科博士課程修了 (自然言語処理学専攻)
1992年 北陸先端科学技術大学院大学助教授	博士 (工学)
2000年 東京工業大学助教授	2003年 東京工業大学助手
2007年 東京工業大学准教授	2007年 東京工業大学助教
2009年 東京工業大学教授	2010年 東京工業大学准教授
現在に至る	現在に至る

言語処理のための機械学習入門

Introduction to Machine Learning for Natural Language Processing

© Hiroya Takamura 2010

2010年8月5日 初版第1刷発行

検印省略

監修者 奥村 学
著者 高村 大也
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 978-4-339-02751-8 (河村) (製本: 愛千製本所)

Printed in Japan



無断複写・転載を禁ずる

落丁・乱丁本はお取替えいたします