

まえがき

本書は、「意味」という抽象的な概念を数学的に記述するために発展した「統計的潜在意味解析」という研究分野に関して体系的にまとめた専門書である。私の知る限り、洋書を含めて、この分野に関して本書ほど詳細に説明したものはないと自負している。

「なぜ「意味」という抽象的な概念を数学的に記述したいのか」というと、それは機械が「意味」を扱えるようにしたいからである。つまり、この分野は、機械学習の一分野として発展してきた。特に、統計モデルによって記述することを目標とする。「なぜ統計モデルによって記述したいのか」というと、それは可能な限り人間の恣意性を排除し、「データ」から「意味」を表現する情報を取り出したいからである。じつは、ここで扱う「意味」の解析というのは言語処理の分野に留まらない。例えば、画像、音声、ゲノム、人の購買履歴などの応用では、それらのデータに対して「単語」に相当する情報を表現することによって、その「意味」の解析が言語処理の場合と同様に可能となってきた。そもそも、言語データに現れる「単語」も情報の一つの表現にすぎないのである。したがって、本書で扱う内容は、言語学における「意味論」とはまったく異なることに注意が必要である。

本書は、『自然言語処理シリーズ』の中で、おそらく最も機械学習色の強い本である。その理由として、本書が扱っている「トピックモデルによる統計的潜在意味解析」という学問分野が、多様な機械学習技術の上に構築されているからである。したがって、本書を読み進めていくと、数理的な内容が中心となっていくことに気がつくだろう。さらに、内容としてさまざまな数理を扱っていることに初めは戸惑うかもしれない。しかし、ここで扱われている内容は、どれもこの分野では重要なテーマである。

この分野を習得するためには、以下の三要素:

- 確率・統計、最適化に関する基礎力
- 現象を統計モデルによって記述する表現力
- 機械学習技術に関する広範囲な知識

が必要であると筆者は考えている。

そこで本書では、単に「トピックモデルによる統計的潜在意味解析」という一分野の解説をするのではなく、この分野を取り巻くさまざまな機械学習技術に関して可能な限り解説することで、これらの3要素が自然と習得できるようにテーマを選んで構成している。

思えば、ここで必要となる3要素は、自然言語処理に限らずさまざまな応用分野に共通して必要なものである。したがって、本書は、データ解析を必要とするさまざまな方々にとっても有用なものとなるであろう。

最後に、さまざまな方々のご協力を感謝致します。構成の段階から適切なコメントを頂いた東京工業大学の奥村学教授、高村大也准教授に心より御礼申し上げます。お茶の水女子大学の小林一郎教授には、研究室所属の情報科学コース院生との草稿チェックのゼミを設けて頂き、誤記・不正確・不明瞭な点や、可読性のためのフィードバックを多数頂きました。小林教授を始めゼミ参加者の皆様、特に江里口瑛子氏、鈴木聡子氏と、数式のチェックにご協力頂いた金子晃名誉教授に感謝致します。本書を執筆するために必要であった知識は、すべて私の博士課程時代に培われたものです。指導教員である中川裕志教授(東京大学)に、この場をお借りして御礼申し上げます。また、中川研究室の学生との機械学習に関する日々のディスカッションやフィードバックは、私の知識の幅を広げる助けとなりました。学生の皆様に深く感謝致します。コロナ社の編集者の方々には、原稿の遅れのために数々のご迷惑をおかけしたにもかかわらず、終始温かく見守って頂きました。改めまして皆様に厚く御礼申し上げます。

2015年2月

佐藤 一誠

目 次

0. 本書の使い方

0.1 本書の読み方	1
0.2 各章と付録の説明	2
0.3 本書で用いる記号など	4

1. 統計的潜在意味解析とは

1.1 潜在的意味・トピックと潜在的共起性	8
1.2 潜在意味解析の歴史	10
1.3 統計的潜在意味解析とデータ駆動インテリジェンスの創発	14
1.4 確率的潜在変数モデル	16
1.5 確率的生成モデルとグラフィカルモデル	17

2. Latent Dirichlet Allocation

2.1 概 要	25
2.2 多項分布と Dirichlet 分布	26
2.3 LDA の生成過程	31
2.4 LDA の幾何学的解釈	35
2.5 LDA の応用例	36

3. 学習アルゴリズム

3.1 統計的学習アルゴリズム	40
3.2 サンプルング近似法	45
3.2.1 ギブスサンプリング	45
3.2.2 周辺化ギブスサンプリング	49
3.2.3 LDA のギブスサンプリング	51
3.2.4 LDA の周辺化ギブスサンプリング	55
3.3 変分近似法	58
3.3.1 変分法	58
3.3.2 変分ベイズ法 (1)	59
3.3.3 変分ベイズ法 (2)	70
3.3.4 LDA の変分ベイズ法 (準備)	71
3.3.5 LDA の変分ベイズ法 (1)	73
3.3.6 LDA の変分ベイズ法 (2)	80
3.3.7 LDA の変分ベイズ法 (3)	83
3.3.8 LDA の周辺化変分ベイズ法	84
3.4 逐次ベイズ学習——変分近似法の場合——	90
3.4.1 確率的最適化と逐次学習	91
3.4.2 自然勾配法	93
3.4.3 LDA の確率の変分ベイズ法	96
3.5 逐次ベイズ学習——サンプルング近似法の場合——	99
3.5.1 粒子フィルタ	100
3.5.2 LDA の粒子フィルタ	105
3.6 Dirichlet 分布のパラメータ推定	110
3.6.1 対称/非対称 Dirichlet 分布の性質	110

3.6.2	変分ベイズ法における Dirichlet 分布のパラメータ推定	112
3.6.3	固定点反復法	112
3.6.4	ニュートン・ラフソン法	115
3.6.5	逐次学習-確率的ニュートン・ラフソン法	118
3.6.6	周辺化ギブスサンプリング/変分ベイズ法の場合	119
3.7	評価方法	123
3.7.1	Perplexity	123
3.7.2	各種学習アルゴリズムにおける Perplexity の計算方法	126
3.7.3	新規文書における Perplexity の計算方法	129
3.7.4	Coherence	130
3.8	各種学習アルゴリズムの比較	132
3.9	モデル選択	135

4. 潜在意味空間における回帰と識別

4.1	背景	137
4.2	潜在意味空間における回帰問題	138
4.2.1	正規回帰モデル	138
4.2.2	LDA+正規回帰モデル	139
4.2.3	LDA+正規回帰モデルの学習アルゴリズム	140
4.2.4	一般化線形モデル	145
4.2.5	LDA+一般化線形モデル	147
4.2.6	LDA+ポアソン回帰モデル	147
4.3	潜在意味空間における分類問題	150
4.3.1	LDA+ロジスティック回帰モデル	150
4.3.2	LDA+多クラスロジスティック回帰モデル	153
4.3.3	LDA+SVM	155

4.3.4	LDA+SVM の学習アルゴリズム	158
-------	-------------------	-----

5. 拡張モデル

5.1	相関構造のモデリング	163
5.1.1	モデリング	164
5.1.2	学習アルゴリズム	165
5.2	系列データのモデリング——統語構造のモデリング——	172
5.2.1	モデリング	173
5.2.2	学習アルゴリズム	176
5.3	時系列データのモデリング	184
5.3.1	モデリング	184
5.3.2	カルマンフィルタの基礎	187
5.3.3	学習アルゴリズム	204
5.4	補助情報を考慮したモデリング	213
5.4.1	モデリング	213
5.4.2	学習アルゴリズム	215

付	録	217
---	---	-----

A.1	対数周辺尤度, KL 情報量, 変分下限の関係の証明	217
A.2	ヘッセ行列の逆行列	217
A.3	KL 情報量とフィッシャー情報行列の関係	218
A.4	制約なし最適化	220
A.4.1	最急降下法	221
A.4.2	共役性	222
A.4.3	共役勾配法	222
A.5	制約付き最適化	224

A.5.1	ラグランジュ未定乗数法と KKT 条件	225
A.5.2	ラグランジュ緩和と双対問題	226
A.6	双対座標降下法による SVM の学習	228
A.6.1	SVM の双対問題	228
A.6.2	双対座標降下法	229
A.6.3	双対座標降下法の収束定理	232
A.7	確率的最適化	233
A.8	ラプラス変分ベイズ法	236
A.8.1	ラプラス近似と変分ベイズ法への適用	236
A.8.2	CTM におけるラプラス変分ベイズ法	238
A.9	ベクトルと行列の基本的な演算公式	240
A.10	期待値, 分散, 共分散の基礎	241
A.11	正規分布の基礎	243
A.11.1	対 称 性	243
A.11.2	正規分布間の基本的な計算	243
A.11.3	多次元正規分布の条件付き確率分布	247
	引用・参考文献	251
	索 引	257



本書の使い方

本章では、本書の読み方および本書で用いる記号について説明する。なお、本書を読む前に本シリーズの『言語処理のための機械学習入門』⁵³⁾†をあらかじめ読んでおくことが望ましい。もしくは、必要に応じて関連項目を参照されたい。

0.1 本書の読み方

図 0.1 に、各章と付録の関係を示す。基本的な読み方としては、1 章から読み進め、付録を適宜参照するという直線的な方法をお勧めする。これは、図中の「一般的な読み方」を表す矢印の順番通りに読み進めていけばよい。

各章から付録への矢印には 2 種類ある。実線の矢印は、各章の中で証明などを省くときに用いる。特に、証明は後で確認すればよいという方は、省略して構わないが、後回しにしたとしてもいずれは読んでおいたほうがよいだろう。一点鎖線の矢印は、発展的な内容になるため、意欲のある読者のみを対象としている。特に、現時点における機械学習技術について可能な限り新しい技術について取り上げている。それゆえに、数学的には難しい内容となっているところもある。

付録から各章への矢印は、事前知っておいたほうがよい基礎知識である。依存関係のある各章へ入る前に事前に読んでおくとういだろう。

図中の「概要をつかむ読み方」は、モデリングだけを先取りして概観すると

† 肩付き数字は、巻末の引用・参考文献の番号を表す。

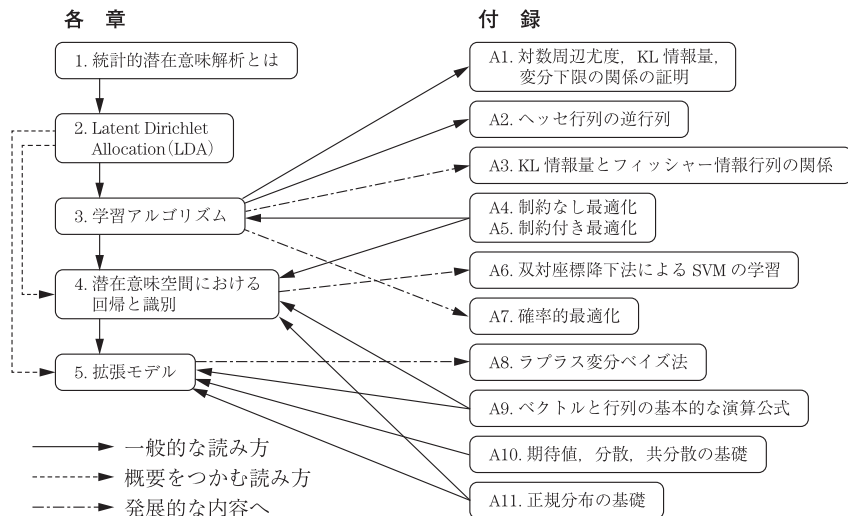


図 0.1 各章と付録の関係および本書の流れ

いう読み方である。本書が数学的に難しいと感じる原因は、学習アルゴリズムの導出にある。したがって、学習アルゴリズムの導出はひとまずおいて、モデリングの部分だけを読み進めてもよいだろう。このように概観することで、本書でどのような問題とモデリングを習得できるのかわかるはずである。

0.2 各章と付録の説明

1章では、統計的潜在意味解析に関する概要と歴史的な背景を説明する。「潜在的意味」や「潜在トピック」についての説明[†]を行う。また、統計的潜在意味解析が基礎として用いる確率的潜在変数モデルについて説明する。「潜在的意味」や「潜在トピック」というのは、一言でいえば、確率的潜在変数モデルにおける潜在変数が示す状態のことである。また、データの生成過程に着目した確率的生成モデルという考え方や、統計モデルを視覚的に表現するグラフィカルモデルについて説明する。

[†] ただし、明確な定義があるわけではない。

2章では、現在この分野での中心的な確率的潜在変数モデルである Latent Dirichlet Allocation (LDA) について概要を説明する。また、LDA の言語データ以外への応用についても紹介する。

3章では、LDA の学習アルゴリズムについて説明する。機械学習の研究は、おもに以下の二つからなる。

- モデリングの研究：対象をどのようにモデリングするか
- アルゴリズムの研究：あるモデルをどのように学習するか

LDA 自体は、統計モデルにすぎず、2章では、最初の「文書をどのようにモデリングするか」というのがテーマである。本章では、二つ目の「LDA をどのように学習するか」について説明する。これまで、LDA の学習アルゴリズムとして多数の方法が提案されてきた。本書では、この学習アルゴリズムの代表的なものを紹介する。おそらく、この3章が本書の最初の「山場」である。多くの読者が本書で挫折するとすれば、この章である。しかし、本書の特徴の一つはこの3章にあるとあってよい。ここでは、LDA の学習アルゴリズムの導出を目標として、近年の機械学習分野で提案されているさまざまな学習アルゴリズムを説明する。したがって、本章を習得すれば、確率的潜在変数モデルなどで近年よく使われている学習アルゴリズムを一気に習得できてしまうのである。

4章と5章では、LDA の拡張モデルに関する研究について説明する。これらの章の目的は、おもに三つある。

- 3章で身につけた学習アルゴリズムの知識を発展させる：3章で学ぶ学習アルゴリズムだけでは、多種多様な LDA の拡張モデルに対応できない。したがって、本章で用いられている数学的な技術を習得することで、より広範囲なモデルに対する学習アルゴリズムの導出が可能となる。
- ほかの機械学習技術の知識を身につける：LDA の拡張モデルは、ほかの機械学習技術と結びついているため、それらの知識を同時に身に付ける。例えば、機械学習技術として有名な、線形回帰モデル（さらに一般化線形モデルへの拡張）、Support Vector Machine (SVM)、ロジスティック回帰モデル、隠れマルコフモデル、カルマンフィルタなどについて知る

4 0. 本書の使い方

ことができる。

- 現象をモデリングする技術を習得する：LDA を拡張する研究では、データに現れるであろうさまざまな現象を統計モデルで表現する。したがって、LDA の拡張モデルを学ぶことは、現象を統計モデルで表現する方法を学ぶことにほかならない。

最後に、付録には、本書で必要となる確率・統計、最適化の基本的な項目を載せた。さらに、発展的な内容として、近年の機械学習技術の中で重要と思える項目を入れた。例えば、SVM の学習アルゴリズムは、Sequential Minimal Optimization (SMO) が有名であるが、ここでは、比較的新しい双対座標降下法と呼ばれる方法を説明した。

また、付録 A.11.3 で説明している「多次元正規分布の条件付き確率分布」の計算のように比較的有名な計算方法には、従来がよく知られた導出（この場合は逆行列を駆使するものがよく知られている）ではなく、多少視点の異なる導出を説明することで、ほかの本とは異なる知識が得られるように工夫している。したがって、すでに知っている項目も一度目を通していただきたい。

0.3 本書で用いる記号など

本書を読み進めていくと

————— 以下を読む前に考えてみよう —————

という箇所が出てくるだろう。ここでは文字とおりに、一度立ち止まって自分で考えていただきたい。本書には、練習問題を挙げない代わりに、自分で考えるタイミングを用意した。研究で重要なことは、自分の頭で考えることである。手を動かさなくとも一度立ち止まって考えてみるだけでもよい。ここで重要なことは、練習問題として自分で「解く」という行為ではなく、自分自身で「考える」ことなのであるから、それを意識してもらいたい。

式中の 波線や 二重線 は、その箇所を注目して欲しい場合に用いている。特

に、式変形の補助的な役割の場合が多い。例えば、式中で波線が付けられた項は、近くにあるほかの項に付けられた波線に関係している、もしくは同一の式の場合がある。

本書の式には、一部を除き原則として式番号を付けた。通常、参照しない式には式番号を付ける必要はないが、議論の場では「式 (29) が」などと直接指すことができるかと話が早いからである。

そのほかいくつかの基本的な記号について説明しておく。ここでは、本書で使わなかったとしても、本書に関係する論文などを読み進める上でよく使われる記号についても説明しておく。また、本書中でも、適宜説明は入れるが、わからない記号などがあれば、ここへ戻ってきて参照されたい。

- \mathbb{R}^K : K 次元実数ベクトルの集合を表す。
- 集合の記号 : 集合 $\{x_1, x_2, \dots, x_n\}$ を $\{x_i\}_{i=1}^n$ などと表す。
- バックスラッシュ「\」 : 集合から特定の集合を取り除くときに用いる。例えば、集合 $\{a, b, c, d\}$ から集合 $\{a, b\}$ を取り除いた集合を $\{a, b, c, d\} \setminus \{a, b\}$ ($= \{c, d\}$) などと書く。また、要素を取り除く場合は、簡略化してバックスラッシュの後に要素を表す記号のみ用いる。例えば、 $\{b, c, d\} = \{a, b, c, d\} \setminus a$ などと書く。さらに、集合によって計算される量から特定の要素を省く場合にもバックスラッシュを用いる。例えば、 $\{x_i\}_{i=1}^n$ の各要素には 1 から 6 の値が入っているとしよう (サイコロの出た目を想像するとよい)。 $\{x_i\}_{i=1}^n$ の中で値が k であるような要素の個数を n_k とする[†]。 j 番目の要素の値 x_j を省いた個数を $n_k^{\setminus j}$ などと上付き添字で表現する。
- 太字 : ある変数のベクトルは太字で表す。例えば、 $\mathbf{x} \in \mathbb{R}^n$ は、 $\mathbf{x} = (x_1, x_2, \dots, x_n)$ 、 $x_i \in \mathbb{R} (i = 1, 2, \dots, n)$ である。また、変数 $x_{i,j} (i = 1, 2, \dots, n, j = 1, 2, \dots, m)$ のとき、片方の添字を伏せた太字は、 $\mathbf{x}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,m})$ を表し、すべての添字を伏せた太字は、 $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ を表す。

[†] もちろん、集合なので値だけみると同じ値のものは 2 個以上存在しない。ここでは、変数 x_i の集合であり、変数としてはそれぞれ違うものであると考えていただきたい。

6 0. 本書の使い方

- $\mathbf{0}$: 要素がすべて 0 のベクトルを表す。
- 和記号 : $i = 1, 2, \dots, n$ のとき, $\sum_{i=1}^n$ を省略して, \sum_i などと書く。また, 集合 $S = 1, 2, \dots, n$ を用いて, $\sum_{i \in S}$ などと書く場合もある。 j 以外の和を取る場合は, $\sum_{\{1, 2, \dots, n\} \setminus j}$ または $\sum_{i \neq j}^n$ などと書く。確率変数 z が $1, 2, \dots, K$ の値を取り, それぞれの確率を $p(z = k)$ とするとき, ある z の関数 $f(z)$ の $p(z)$ による期待値計算を $\mathbb{E}_{p(z)}[f(z)] = \sum_{k=1}^K q(z = k) f(z = k) = \sum_z q(z) f(z)$ などと省略して書く。
- ドット「 \cdot 」による和 : 変数 $x_{i,j}$ ($i = 1, 2, \dots, n, j = 1, 2, \dots, m$) のとき, $x_{i,\cdot} = \sum_{j=1}^m x_{i,j}$ および $x_{\cdot,j} = \sum_{i=1}^n x_{i,j}$ を表す。
- O : 要素がすべて 0 の行列を表す。
- $\text{Tr}(A)$: 行列 A のトレースを表す。
- $\text{diag}(\mathbf{v})$: ベクトル $\mathbf{v} = (v_1, \dots, v_K)$ を対角要素とする対角行列を表す。
- $|A|$: 行列 A の行列式を表す。
- A^\top : 行列 A の転置を表す。
- A^{-1} : 行列 A の逆行列を表す。
- I または E : 単位行列を表す。式の見やすさによって使い分ける。
- $A \equiv B$: 「 A を B と定義する」を意味する。
- $x \perp\!\!\!\perp y | z$: z が与えられた下で, x と y は条件付き独立性である。
- $\int f(x, y) dx dy$: \int で一重積分と多重積分の両方を表す。
- 偏微分 : 関数 $f(x, y)$ の, x に関する偏微分を $\frac{\partial}{\partial x} f(x, y)$ または $\frac{\partial f(x, y)}{\partial x}$ と書く。これを単に「微分する」と書くこともある。
- $x \sim P(x)$: 「確率分布 $P(x)$ から x が生成される」, または「確率変数 x が確率分布 $P(x)$ に従う」ことを表す。例えば, 平均 m , 分散 σ の 1 次元正規分布を $\mathcal{N}(x|m, \sigma)$ と表したとき, $x \sim \mathcal{N}(x|m, \sigma)$ として「 $\mathcal{N}(x|m, \sigma)$ 」

から x が生成される」, または「確率変数 x が $\mathcal{N}(x|m, \sigma)$ に従う」ことを表す。 $x \sim \mathcal{N}(\underline{m}, \sigma)$ のように x を省略して書くこともある。

- $\delta(\text{条件式})$ または $\mathbf{1}(\text{条件式})$: 条件式が満たされたとき 1 を, そうでないときに 0 を返す関数である。例えば, $\delta(x = y)$ は $x = y$ のときに 1 を返す。 $\delta(\cdot)$ と $\mathbf{1}(\cdot)$ は, 式の見やすさによって使い分ける。ちなみに, 括弧 $()$ 内の変数を省略して関数を書く場合には, このようにドットを使うこともある。

索引

<p style="text-align: center;">【い】</p> <p>イエンセンの不等式 61</p> <p>一括学習 90</p> <p>一般化線形モデル 145</p> <p>因子分解 60</p> <p style="text-align: center;">【お】</p> <p>オンライン学習 91</p> <p style="text-align: center;">【か】</p> <p>回帰問題 137</p> <p>階層ベイズモデル 14</p> <p>ガウス回帰モデル 139</p> <p>ガウス分布 139</p> <p>確率的勾配 92</p> <p>確率的最適化 91</p> <p>確率的生成モデル 17</p> <p>確率的潜在変数モデル 16</p> <p>確率的変分ベイズ法 91</p> <p>隠れマルコフモデル 172</p> <p>カルバック・ライブラー 情報量 41</p> <p>ガンマ関数 27</p> <p style="text-align: center;">【き】</p> <p>ギブスサンプリング 45</p> <p>教師あり学習 137</p> <p>強凸関数 233</p> <p>共役勾配法 149</p> <p>共役事前分布 28</p> <p>近似事後分布 60</p> <p style="text-align: center;">【く】</p> <p>グラフィカルモデル 17</p>	<p style="text-align: center;">【け】</p> <p>系列データ 172</p> <p style="text-align: center;">【こ】</p> <p>固定点反復法 112</p> <p style="text-align: center;">【さ】</p> <p>最急降下法 220</p> <p>最適解の必要条件 221</p> <p>最適性条件 221</p> <p>最尤推定 42</p> <p>座標降下法 229</p> <p>サンプリング近似法 45</p> <p style="text-align: center;">【し】</p> <p>識別問題 137</p> <p>事後確率最大推定 43</p> <p>自己相互情報量 131</p> <p>自然勾配法 93</p> <p>弱双対定理 227</p> <p>周辺化 49</p> <p>周辺化ギブスサンプリング 45</p> <p>周辺化変分ベイズ法 84</p> <p>主問題 227</p> <p>条件付き独立性 20</p> <p>状態空間モデル 185</p> <p>人工知能 14</p> <p>シンタックス 173</p> <p style="text-align: center;">【せ】</p> <p>正規回帰モデル 139</p> <p>正定値 222</p> <p>正定値行列 221</p> <p>制約付き最適化問題 224</p> <p>制約付きベイズ学習 157</p>	<p>制約なし最適化問題 220</p> <p>積分消去 49</p> <p>線形回帰モデル 137</p> <p>潜在トピック 9</p> <p>潜在トピックモデル 25</p> <p>潜在変数 16</p> <p style="text-align: center;">【そ】</p> <p>双対関数 227</p> <p>双対ギャップ 227</p> <p>双対座標降下法 229</p> <p>双対問題 227</p> <p>ソフトマージン 155</p> <p style="text-align: center;">【た】</p> <p>多クラスロジスティック 回帰モデル 153</p> <p>多項分布 26</p> <p>単体 27, 35</p> <p style="text-align: center;">【ち】</p> <p>逐次学習 91</p> <p>直線探索 221</p> <p style="text-align: center;">【て】</p> <p>ディガンマ関数 71</p> <p>テイラー展開 87</p> <p>低ランク近似行列 11</p> <p>停留点 221</p> <p>データ駆動 インテリジェンス 14</p> <p>点推定 43, 83</p> <p style="text-align: center;">【と】</p> <p>統計的潜在意味解析 8</p> <p>統語論 173</p>
--	--	---

等式制約付き最適化問題	225
特異値分解	10
トライガンマ関数	117
【に】	
ニュートン・ラフソン法	112
人間の知能	14
【は】	
破棄する期間	51
バッチ学習	90
汎化能力	42
【ふ】	
フィッシャー情報行列	95
フィルトレーション	234
ブサイ関数	71
不等式制約付き最適化問題	225

【へ】	
ベイズ学習	20
ベイズの定理	20
ヘッセ行列	116
変分下限	61
変分ベイズ法	58
変分法	58
【ほ】	
ポアソン回帰モデル	146
ポアソン分布	146
補助情報	213
ポラック・リビエール法	224
【ゆ】	
優マルチンゲール収束定理	233

【ら】	
ラグランジュ関数	225
ラグランジュ緩和	226
ラグランジュ乗数	225
ラグランジュ未定乗数法	225
ラプラス近似	236
ラプラス変分ベイズ法	170, 236
【り】	
リサンプリング	104
リプシッツ条件	234
粒子フィルタ	99
【ろ】	
ロジスティック回帰モデル	137
ロビンズ・モンロー型	91

【A】	
artificial intelligence	14
【B】	
bag of words	25
batch learning	90
Bayesian learning	20
Bayes theorem	20
BoW	25
burn-in period	51
【C】	
classification problem	137
Coherence	123
collapsed Gibbs sampling	45
collapsed variational Bayesian method	84
conditional independence	20

conjugate gradient method	149
conjugate prior	28
constrained Bayesian learning	157
coordinate descent method	229
correlated topic model	163
C _T M	163
CVB	84
【D】	
data-driven intelligence	14
Dirichlet 分布	26
Dirichlet distribution	26
Dirichlet-多項回帰モデル	215
Dirichlet-multinomial regression model	215
DTM	184
dual	227
dual coordinate descent	

method	229
dual function	227
duality gap	227
dynamic topic model	184
【E】	
effective sample size	104
ESS	104
【F】	
factorization	60
filtration	234
Fisher information matrix	95
fixed point iteration	112
【G】	
gamma function	27
Gauss distribution	139
Gauss regression model	139
generalization ability	42

generalized linear model	145	Lipchitz condition	234	probabilistic latent variable model	16
Gibbs sampling	45	logistic regression model	137		
graphical model	17	low rank approximation	11	[R]	
[H]		[M]		regression problem	137
Hessian matrix	116	MAP 推定	43	resampling	104
hidden Markov model	172	maximum a posteriori 推定	43	Robbins–Monro 型	91
hierarchical Bayes model	14	maximum likelihood estimation	42	[S]	
HMM	172	multi-class logistic regression model	153	simplex	27, 35
HMM–LDA	172	multinomial distribution	26	singular value decomposition	10
human intelligence	14			soft margin	155
[J]		[N]		state space model	185
Jensen の不等式	61	natural gradient method	93	stationary point	221
[K]		Newton–Raphson 法	112	statistical latent semantic analysis	8
Karush–Kuhn–Tucker 条件	226	normal regression model	139	steepest descent method	220
KKT	226			stochastic gradient	92
KL divergence	41	[O]		stochastic optimization	91
Kullback–Leibler 情報量	41	online learning	91	stochastic variational Bayesian method	91
[L]		[P]		strongly convex function	233
Lagrange function	225	particle filter	99	supervised learning	137
Lagrange multiplier	225	Perplexity	123	support vector machine	137
Lagrange multiplier method	225	PMI	131	SVM	137, 155
Lagrange relaxation	226	point estimation	43, 83	syntax	173
Laplace approximation	236	pointwise mutual information	131	[T]	
Laplace variational Bayesian method	170, 236	Poisson distribution	146	Taylor expansion	87
latent Dirichlet allocation	25	Poisson regression model	146	trigamma function	117
latent topic	9	Polak–Ribiere method	224	[V]	
latent topic model	25	Positive-definite	222	variational Bayesian method	58
latent variable	16	positive-definite matrix	221	variational lower-bound	61
LDA	25	primal	227	variational method	58
linear regression model	137	probabilistic generative model	17	VB	58
line search	221				

— 監修者・著者略歴 —

奥村 学 (おくむら まなぶ)	佐藤 一誠 (さとう いっせい)
1984年 東京工業大学工学部情報工学科卒業	2011年 東京大学大学院情報理工学系研究科博士課程修了 (数理情報学専攻), 博士 (情報理工学)
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	2011年 東京大学情報基盤センター助教
1989年 東京工業大学助手	2013年 科学技術振興機構 さきがけ研究員 (兼務)
1992年 北陸先端科学技術大学院大学助教授	現在に至る
2000年 東京工業大学助教授	
2007年 東京工業大学准教授	
2009年 東京工業大学教授	
現在に至る	

トピックモデルによる統計的潜在意味解析

Statistical Latent Semantic Analysis Based on Topic Model

© Issei Sato 2015

2015年4月6日 初版第1刷発行

検印省略

監修者 奥村 学
著者 佐藤 一誠
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 978-4-339-02758-7 (新井) (製本: 愛千製本所)

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おります。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