

# 統計学の使い方より、本質を見抜く力

## —機械学習の数理100問シリーズと凸最適化への期待—

鈴木 譲

データサイエンスは、データの背後にある本質を見出す営みである。大学教員でもある筆者は、学生が知識の習得を目的に大学で勉強することに、強い疑念を抱いている。データサイエンスに限らず、必要な知識は毎日変化している。また、インターネットがあるので、知識は必要なときにいつでも習得できる。そして、先生の話や本に書いてあることでも、今日は真であったとしても、明日は真である保証はない。大学時代に得られた知識を固定してしまっただけで、卒業してから、生きた化石になるのではないかと、先生の話やインターネットにあることについて、真偽をその都度チェックし、本質を見抜く力をつけたい。統計学の知識の習得は、そのような姿勢を身に着けた後でも遅くはないのではないかと。

キーワード：データサイエンティスト、統計学、スパース推定、凸最適化

### 1. まえがき

今回、本誌から、特集「データサイエンス」で何か記事を書くよう、声をかけていただいた。私は、早稲田大学理工学部工業経営学科の出身で、当時森戸晋先生が「整数計画法と組合せ最適化」(今野浩・鈴木久敏著、日科技連) [1] というテキストを使って大学院の講義(通年)をされていて、単位をとった記憶がある。それも手伝ってか、大昔のことだが、分枝限定法を用いたベイジアンネットワークの構造学習の効率化(無駄な探索を枝刈りする)という論文を書いて、機械学習の最高峰として知られるICMLという会議で発表したことがある [2]。OR から離れて久しいが、異なる視点から話題を提供できればと思います、筆を取らせていただくことになった。

最近のデータサイエンスブームで、国内でもデータサイエンス学部や関連した学科・コースが誕生している。そして、統計学に関する知識を習得し、生涯データサイエンスの業界でやっていくプランをもつ学生も多くなっている。これは、データサイエンスの指導者にとって、大変喜ばしいことである。

まず、データサイエンスは、データの背後にある本質を見出す営みである。統計学は、パラメータ推定や統計検定のように、データを生成しているメカニズムを発見するための方法論である。

かれこれ20年ほど前になるが、柴田里程先生が「デー

タサイエンスのすすめ」という記事を書かれていて [3]、それが日本のデータサイエンスの始まりとされている。統計学は、昔からの理論や手法を議論する場になっていて、本来の目的から乖離しているという。そもそも、統計学やORといった実用的な学問は、その目的は同じであっても、何を議論すべきかは時代とともに変化すべきものであると。

筆者は、当時そのような問題意識は理解できるが、具体的にどのように実現していくかが見えなかった。3年前に現在の職場(阪大基礎工学数理)に異動になって、統計学に深く関係するようになってから、その件について、あることを実践してきた。タイトルにもあるような「統計学の使い方より、本質を見抜く力」という視点に立って、数学やプログラミングを中心に学生のロジックを伸ばす教育を行っている。知識は意味がないというのではなく、後からついてくるものという認識をしている。最近「機械学習の数理100問シリーズ」(共立出版)という書籍を出版し、その視点をアピールしている。その周辺の話題を提供することが本稿の目的である。

また、研究の分野でも、統計学だけではなく、機械学習や最適化理論などの境界領域を含めた複合領域で研究すれば、よいテーマが見つかりやすい、よりグローバルな研究結果が出ているという視点をもっている。今回、書籍の執筆にあたって調査した結果、統計学でよく用いられているスパース推定などで、凸最適化の手法が非常によく用いられていることがわかった。最後に、そのあたりの事例を紹介する。

すずき じょう

大阪大学大学院基礎工学研究科

〒560-8531 大阪府豊中市待兼山1-3

j-suzuki@sigmath.es.osaka-u.ac.jp

## 2. データサイエンス教育をとりまく環境と問題の所在

### 2.1 使いだけを教育することの危うさ

統計学を教えていると、理屈はいつでもよいから、使い方を教えて下さいという人がある。そういう人に限って、状況ごとに判断するのではなく、〇〇という状況では〇〇検定を適用し、R言語の〇〇というパッケージにデータを入力して出力を見るというように、丸暗記の対応をしている。したがって、結果の解釈が表面的であったり、不具合にも対応できなかつたりする。そして、統計学の方法は、各時点で最適なものが得られているとは限らず、時を追うごとに少しずつ改善している。そのような状況と行動の対応は、学生時代に覚えたものを生涯にわたって使うのではなく、徐々に更新していくものである。変化に追従できるよう、本質を学んでおく必要がある。

また、最近では、たとえばIBMのWatson Studioのように<sup>1</sup>、Auto AIといって、条件を入れるとデータ処理のモデルを作成して、分析してくれるようなソフトウェアも多数出てきている。つまり、データサイエンスの業務は、統計学の使いだけを覚えればよいという消極的な態度では、到底おぼつかない。

### 2.2 いわゆる統計学だけでよいのか

最近では、インターネットの普及により、大学以外でも、オンライン学習など、効率よく統計学を学ぶ手段がいくつもできている。統計学に限らず、知識を得るための手段が増えてきて、バブル時代には、品質管理や統計関連の高価なセミナーが繁盛していた。しかし現在では、インターネットの普及で、そのようなビジネスは過当競争になっている。

そもそも、データサイエンスは、物理や生物のような、純粋なサイエンスに近いという考え方もできる。純粋理論でない限り、何かを主張する以上、実験データが必要になる。それを分析して背後にある法則を得るのが、サイエンスの営みである。実際、データサイエンスの会社に行ってみると、博士取得者が多く、しかも統計学以外のサイエンス経験者が多いことに気づく。そうした職場の人に聞くと、学生時代にサイエンスで鍛えてきた本質を理解する能力、学会やジャーナルなどで論文をプレゼンする能力が活かされているという。

データサイエンスにおいて、統計学が重要な位置を占めるということについては、誰も異論がない。しか

し、統計学だけを本質まで理解したとしても、それはデータサイエンスの方法論の一つにすぎない。それだけで広大なデータサイエンスという対象のすべてをカバーしていると言えるだろうか。

### 2.3 データサイエンティスト 25万人、頭脳になる一握りと手足になる大多数になるという現実

データサイエンス従事者が25万人不足しているということが叫ばれ、今後データサイエンスの業務に従事する人が増えてくるであろう。しかし、人員が充足したとして、彼らがすべてハッピーな人生を送れるだろうか。データサイエンス協会では、スキルチェックリスト [4] なるものを提供している。データサイエンス従事者に求められるスキルが、業界を代表するレベル、棟梁レベル、独り立ちレベル、見習いレベルに別れて定義されている。しかし、そのスキルを得るために必要な資質のようなものは、記載されていない。実際に、Aさんなら得られるが、Bさんでは無理というようなスキルが多い。チェックリストを眺めてみて、足りないスキルを追加すればよい、というのではなく、その前に何をすべきかを考える場合が多い。

現在は、頭脳として働いているデータサイエンティストは比較的多い。しかし、業界が大きくなれば、必要な資質をもたないために、棟梁はおろか、独り立ちにも達しないレベルのデータサイエンティストが大多数になることが予想される。

データサイエンスの仕事から得られる満足、魅力のようなものはいくつもある。たとえば、7名のチームのうち、他の6名が気づかなかった本質を自分だけが見出したといった、いわゆる存在感を示すことなどもその一つだ。しかし、逆にそういうことがないと、年功序列や終身雇用の崩壊した今日、「いてもいなくても同じ」「代わりはたくさんいる」という人事評価になり、契約期間を延長しない、早期退職の候補になるということもありうる。いずれにせよ、頭脳にならないと、データサイエンス本来の満足は得られないのではないか。そして、生き残れないのではないか。

## 3. 統計学の使い方より、本質を見抜く力

それでは、本質を見抜く力は、どうしたら得られるのか。

筆者は、仕事から、学生の就職や共同研究などで、データサイエンス企業との付き合いが多い。活躍している人に聞くと、知識や経験より、ロジックが重要であると答える人が多い。ロジックができてくると、他人が見えない本質やチャンス（自分だけが）見るこ

<sup>1</sup> [https://www.ibm.com/jp-ja/cloud/watson-studio/](https://www.ibm.com/jp-ja/cloud/watson-studio/autoai)  
autoai

ともできるという。知識や経験は得ようと思えばいつでも得られる。知識を得るのに精一杯であれば、それはロジックが足りない証拠であると。

以下、脳裏にロジックを構築する方法として、線形代数とプログラミングについて説明したい。

### 3.1 線形代数

最近、高校の数学で、行列をカリキュラムからはずし、データ分析を教えることに対する批判を、特にデータサイエンス関係者以外からよく聞く。個人的には、データ分析は、情報という科目の中にいれて、行列を復活すべしという考えをもっている。いずれにせよ、線形代数の重要性を主張している。

よく、線形代数を理解していないので、データ分析で疑問点が生じるという話をよく聞く。筆者は、その場合に、線形代数とその周辺がわかっていないのではなく、ロジックそのものができていないという指摘をする。高校時代に理系文系にわかれるのは日本だけのことである。sin, logなどの数学を使わなくても、数学から得られるロジックは、データサイエンスに限らず知的な業務では、強く要求される。また、受験科目に数学のない大学では、入学者は数学ができないという前提で講義をするので、余計差が開いてくる。

もし、まっとうなデータサイエンス人材を育てたいのなら、受験科目に数学を入れて、大学1年生で線形代数の講義を行うべきだ。線形代数は、純粋数学、応用数学の基礎であるばかりでなく、行動を支えるロジックの基礎になる。

### 3.2 プログラミング

データサイエンス教育では、C言語やJavaではなく、R言語やPythonが用いられている。その方が、習得する期間が短いので、小回りがきく。

ただ、データサイエンス教育の現場で、出来合いのパッケージやモジュールにデータを入力するといったように、その使い方を覚えるだけにとどまっている光景をよく目にする。スクラッチからプログラミングするという事は、皆無に近い。パッケージにデータを放り込むだけであれば、小学生でもできるかもしれない。

業務であれば、効率の面などから、(信頼できる)パッケージを使うべきだ。しかし、学習の段階では、入力から出力がどのようにして得られるのか、それを実現するソースプログラムを自分の目で確かめる必要がある。

また、R言語やPythonで実行する処理を数学的に解析して、理論的に分析をするような科目が、データサ

イエンス教育でも増えてきている。しかし、数学的に正しく理解がなされたとしても、その理論がプログラムとして正しく実現されているかを確認しないと、せっかくの思考がゴールに達する前で中断されてしまう。

統計処理のソースプログラムを見るという以外に、自分でプログラムを組んで、デバックしたり、ロジックをチェックすることによって、脳が活性化され、判断力や思考力が改善されるという効果がある。

## 4. 機械学習の数理 100 問シリーズ

以前から、2節で述べた問題意識を抱き、3節で述べたロジックの必要性を抱いていた。そして、現在の所属(阪大基礎工数理)に異動になった2017年から、学部3年生対象の機械学習の数理に関する講義の中で、それらを解決する教育を実践するようになった。

筆者は、「機械学習の数理 100 問」という、線形回帰、分類、リサンプリング、情報量基準、スパース推定、非線形回帰、決定木、サポートベクトルマシン(SVM)、教師なし学習といった内容に関する100問の演習問題を作成した(数学の証明が50問程度、プログラミングの問題(穴埋め+実行)が50問程度)。講義では、処理を数学的に説明した後、それを実現するプログラムを自分でスクラッチから書いて、それを見せながら説明を進めていった。現在の所属では、数学や統計学の科目が多く、また、2年生で半期かけてR言語を習得していたので、学生は筆者の厳しい要求にもついていった。

統計学や機械学習の知識も必要だが、それ以上に数学とプログラミングでロジックを鍛え上げ、本質が見えるように変えていくことが重要だと思った。そして、勤務する大学だけではなく、一般に提供しようと思い、書籍、それもシリーズという形で出版することになった。その名も「機械学習の数理 100 問シリーズ」(共立出版)(図1)。R言語とPythonの両方のニーズがあ



図1 2020年3月に出版された拙書、機械学習の数理 100 問シリーズ「統計的機械学習の数理 100 問 with R」(共立出版)。2020年4月には「統計的機械学習の数理 100 問 with Python」(赤い表紙)が出版されている

表1 機械学習の数理 100 問シリーズ (共立出版)

タイトル	R	Python
統計的機械学習の数理 100 問 [5, 6]	○	○
スパース推定 100 問 [7, 8]	○	○
グラフィカルモデルと因果推論 100 問	○	○
機械学習のためのカーネル 100 問	○	○
渡辺澄夫ベイズ理論 100 問	○	○
統計検定 1 級 100 問	○	

るので、最初の 10 巻は 5 種類 (プログラムの部分のみが異なる)、全部で 6 種類の出版になる (表 1)。「統計的機械学習の数理 100 問 with R」[5] と「統計的機械学習の数理 100 問 with Python」[6] はすでに出版されている<sup>2</sup>。また、Springer から、英語版が “Statistical Learning with Math & R” というタイトルで出版される<sup>3</sup> (ebook 版は発行済)。

使い方だけを教える書籍は他人でも書けるし、そのような情報はインターネットでも得られる。つまり cheap な情報のように思えた。だから、そうならないような書籍にした。また、「お話だけで終わらない」ということを心がけた。ほぼすべての説明にソースコードを提供し、git からダウンロードできるようにした<sup>4</sup>。つまり、理論から実現まで、抽象的な段階から具体的な段階まで、思考が中断しないようにした。

また、パッケージがあればソースコードはいらない、という話を聞くことがある。お話や数式だけでは理解できないので、演習として具体的にソースコードを読んでいる。もし、具体的な処理内容がわからず、パッケージの使い方だけを覚えてコマンドを打ち込むだけであれば、前述のように AI に仕事を奪われる可能性すらある。ソースコードをスクラッチのレベルで読み書きすることは、色々な学習効果がある。

## 5. 凸最適化への期待

2.2 節で、統計学だけでよいのか、という疑問を投げた。筆者自身、前節の機械学習の数理の講義や書籍の執筆で、調査をしていくうちに、統計学だけがデータサイエンスを支えているのではないということを実感した。

たとえば、SVM も双対問題に直せば 2 次計画法に帰着でき、カーネルを適用できるようになる [5, 6]。また、最近、機械学習の数理 100 問シリーズのスパース推定を執筆するにあたって調査をしたところ、内部の処

理のほとんどが凸最適化で動いていることがわかった。スパース推定発祥の地とされる Stanford の統計学科の R. Tibshirani, T. Hastie, B. Efron, J. Friedman と著名な教授その周辺では、20 年以上前から、古典的な統計学だけではなく、計算量理論や最適化理論と組み合わせ、有用な問題を定式化しては解くというやりかたをしていることがわかった。筆者も本年度から、新 M1 と Boyd and Vandenberg “Convex Optimization” というテキスト [9] を輪講している。平易であるが、毎回演習問題を 10 問ずつ解いて議論している (全部で 350 問程度ある)。

以下、釈迦に説法かもしれないが、スパース推定の書籍の第 4 章に相当する箇所から、Lasso の凸最適化による解法について紹介してみたい。詳細は、2020 年 10 月出版予定の機械学習の数理 100 問シリーズ「スパース推定 with R」[7] もしくは 11 月出版予定の「スパース推定 with Python」[8] を参照されたい。

Lasso でよく例として用いられているのが、線形回帰の Lasso に関するものである。 $X \in \mathbb{R}^{N \times p}$ ,  $y \in \mathbb{R}^N$ ,  $\lambda > 0$  として、

$$\frac{1}{2} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \quad (1)$$

を最小にする  $\beta \in \mathbb{R}^p$  を求める問題である。第 2 項が L1 ノルムであって、絶対値がある一定値以下になると 0 になる。その意味で、 $p$  変数の係数  $\beta$  のうちの係数が 0 でない成分を見出す処理 (変数選択) になる。

線形回帰の Lasso を変形して、平滑化の目的で、Fused Lasso が用いられている。観測データ  $y = (y_1, \dots, y_N) \in \mathbb{R}^N$  から、

$$\frac{1}{2} \sum_{i=1}^N (y_i - \theta_i)^2 + \lambda \sum_{i=1}^{N-1} |\theta_i - \theta_{i+1}| \quad (2)$$

を最小にする  $\theta = (\theta_1, \dots, \theta_N) \in \mathbb{R}^N$  を求める問題になる。 $y$  を  $\theta$  で平滑化することになる。1 次元での適用が多いが、2 次元であれば、 $(x, y)$  平面で  $z$  方向の値に近い観測データどうし、同一視する処理である。図 2 は、Fused Lasso によって、コロナ感染者数 (2020 年 6 月 11 日現在) が類似の地区同士を同じ色で表示している。

### 5.1 動的計画法による解法

動的計画法を用いると [10], Fused Lasso を以下のよう解くことができる。(2) を最小にする  $\theta_1, \dots, \theta_N$  を解くために、 $\theta_1$  に関する最適条件を求めると、

$$h_1(\theta_1, \theta_2) := \frac{1}{2} (y_1 - \theta_1)^2 + \lambda |\theta_2 - \theta_1|$$

の最小化になるが、 $\theta_2$  が変数として残る。しかし、 $\theta_2$

<sup>2</sup> <https://www.kyoritsu-pub.co.jp/series/214/>

<sup>3</sup> <https://www.springer.com/jp/book/9789811575679>

<sup>4</sup> <https://bitbucket.org/prof-joe>

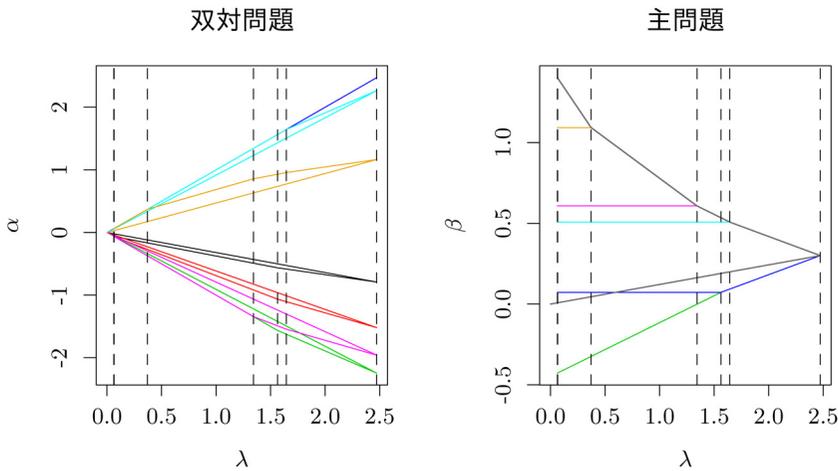


図3  $p = 8, m = 7$  の場合の双対問題の解パス (左) と主問題の解パス (右)  
 行列  $D$  として、1次元の Fused Lasso に相当するものを選んだ。いずれも、 $\lambda$  を下げていくと解がマージしていく。双対問題は  $\alpha \in \mathbb{R}^m$  の解 (7色の折れ線)、主問題は  $\beta \in \mathbb{R}^p$  の解 (8色の折れ線)。

$\lambda = 10$

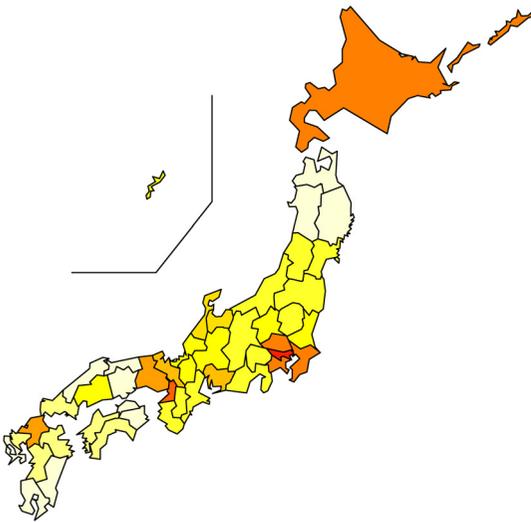


図2 Fused Lasso (2次元の場合) の実例  
 2020年6月11日現在のコロナの感染者数。都道府県の隣接行列の情報から、Fused Lasso で感染者数が近い都道府県どうし、同じ色になるようにした。

の値がわかっている場合の最適な  $\theta_1$  は、

$$\hat{\theta}_1(\theta_2) = \begin{cases} y_1 - \lambda, & y_1 \geq \theta_2 + \lambda \\ \theta_2, & |y_1 - \theta_2| < \lambda \\ y_1 + \lambda, & y_1 \leq \theta_2 - \lambda \end{cases}$$

と書ける。次に、 $\theta_2$  に関する最適な条件を求める場合、

$$\frac{1}{2}(y_1 - \theta_1)^2 + \frac{1}{2}(y_2 - \theta_2)^2 + \lambda|\theta_2 - \theta_1| + \lambda|\theta_3 - \theta_2|$$

の最小化になる。 $\theta_1, \theta_3$  が変数として残るが、 $\theta_1$  を  $\hat{\theta}_1(\theta_2)$  でおきかえると、 $\theta_3$  の値がわかっている場合の  $h_2(\hat{\theta}_1(\theta_2), \theta_2, \theta_3)$  を最小にする  $\theta_2$  の値  $\hat{\theta}_2(\theta_3)$  が  $\theta_3$  の関数として書ける。また、 $\hat{\theta}_1(\theta_2)$  も、 $\theta_3$  の関数  $\hat{\theta}_2(\theta_3)$  として書けるので、 $\hat{\theta}_1(\theta_3)$  と書くことにする。この操作を繰り返していくと、 $\hat{\theta}_1(\theta_N), \dots, \hat{\theta}_{N-1}(\theta_N)$  が  $\theta_N$  の関数として求まり、1変数  $\theta_N$  に関する方程式

$$h_N(\hat{\theta}_1(\theta_N), \dots, \hat{\theta}_{N-1}(\theta_N), \theta_N)$$

を最小にする  $\theta_N$  を求める問題に帰着できる。そして、 $\theta_N$  の値から  $\theta_{N-1} = \hat{\theta}_{N-1}(\theta_N)$  が、 $\theta_{N-1}$  の値から  $\theta_{N-2} = \hat{\theta}_{N-2}(\theta_{N-1})$  が、というように (2) を最小にする  $\theta_1, \dots, \theta_N$  を求めることができる。

### 5.2 Lasso の双対問題による解法

$D \in \mathbb{R}^{m \times p}$  として、

$$\frac{1}{2}\|y - X\theta\|_2^2 + \lambda\|D\theta\|_1 \quad (3)$$

の最小化 (一般化 Lasso [11]) は、 $m = p, D$  が単位行列なら (1) になり、 $m = N - 1, p = N, X$  が単位行列、 $D_{i,j} = 1 (j = i), D_{i,j} = -1 (j = i + 1), 0$  (その他) であれば (2) の最小化になり、

$$\frac{1}{2}\|y - \theta\|_2^2 + \lambda\|\gamma\|_1$$

の  $\lambda = D\theta$  という等式制約の下での  $\theta, \gamma$  に関する最小化とみなせる。Lagrange 乗数  $\alpha$  を導入し、

$$\frac{1}{2}\|y - X\theta\|_2^2 + \lambda\|\gamma\|_1 + \alpha^T(D\theta - \gamma)$$

を  $\theta, \alpha$  で最小化をはかると、 $X^+ := (X^T X)^{-1} X^T$  として

$$\begin{aligned} & \min_{\theta} \left\{ \frac{1}{2}\|y - X\theta\|_2^2 + \alpha^T D\theta \right\} \\ & = \frac{1}{2}\|X X^+ y - (X^+)^T D^T \alpha\|_2^2 \end{aligned} \quad (4)$$

$$\min_{\gamma} \{ \lambda\|\gamma\|_1 - \alpha^T \gamma \} = \begin{cases} 0, & \|\alpha\|_{\infty} \leq \lambda \\ -\infty, & \text{その他} \end{cases}$$

したがって、双対問題は、 $\|\alpha\|_{\infty} \leq \lambda$  のもとで、

$$\frac{1}{2}\|X X^+ y - (X^+)^T D^T \alpha\|_2^2$$

の最小化になる。そして、 $\alpha$  の解  $\hat{\alpha}$  が求まれば、(4) の左辺の最小化される箇所に  $\hat{\alpha}$  を代入して  $\theta$  で微分をすると、 $\hat{\theta} = y - D^T \hat{\alpha}$  となり、 $\hat{\theta}$  の値も求まる。

### 5.3 ADMM による解法

$A \in \mathbb{R}^{d \times m}, B \in \mathbb{R}^{d \times n}, c \in \mathbb{R}^d, f: \mathbb{R}^m \rightarrow \mathbb{R}$  および  $g: \mathbb{R}^n \rightarrow \mathbb{R}$  を凸関数とする。ただし、 $f$  は微分可能であるとする。 $A\alpha + B\beta = c$  のもとで、 $f(\alpha) + g(\beta)$  を最小にする  $\alpha \in \mathbb{R}^m, \beta \in \mathbb{R}^n$  を求める問題を、

$$L(\alpha, \beta, \gamma) := f(\alpha) + g(\beta) + \gamma^T (A\alpha + B\beta - c) \rightarrow \text{最小}$$

( $\gamma \in \mathbb{R}^d$  は未定数) というように Lagrange 未定数法で定式化する。さらに定数  $\rho > 0$  を用いて拡張 Lagrange  $L_{\rho}(\alpha, \beta, \gamma)$  を

$$f(\alpha) + g(\beta) + \gamma^T (A\alpha + B\beta - c) + \frac{\rho}{2}\|A\alpha + B\beta - c\|_2^2 \quad (5)$$

と定義し、 $\alpha_0 \in \mathbb{R}^m, \beta_0 \in \mathbb{R}^n, \gamma_0 \in \mathbb{R}^d$  を適当に決めてから、 $t = 1, 2, \dots$  に対して以下の手順をくりかえす手順 (ADMM, Alternating Direction Method of Multipliers) を適用する。一般化 Lasso の場合、 $L_{\rho}(\alpha, \beta, \gamma)$  を

$$\frac{1}{2}\|y - X\alpha\|_2^2 + \lambda\|\beta\|_1 + \mu^T (D\alpha - \beta) + \frac{\rho}{2}\|D\alpha - \beta\|_2^2$$

とおくと、 $\partial L_{\rho} / \partial \alpha = 0$  および  $\partial L_{\rho} / \partial \beta = 0$  より、更新式は以下ようになる。

$$\begin{cases} \alpha_{t+1} \leftarrow (X + \rho D^T D)^{-1} (y + D^T (\rho \beta_t - \gamma_t)) \\ \beta_{t+1} \leftarrow \mathcal{S}_{\lambda} (\rho D \alpha_{t+1} + \gamma_t) / \rho \\ \gamma_{t+1} \leftarrow \gamma_t + \rho (D \alpha_{t+1} - \beta_{t+1}) \end{cases}$$

ただし、 $A \in \mathbb{R}^{d \times m}, B \in \mathbb{R}^{d \times n}, c \in \mathbb{R}^d, f: \mathbb{R}^m \rightarrow \mathbb{R}, g: \mathbb{R}^n \rightarrow \mathbb{R}$  は、それぞれ  $\frac{1}{2}\|y - X\alpha\|_2^2, \|\beta\|_1, A = D, B = -I, c = 0$  となる。

## 6. むすびに代えて

統計学の使い方より、本質を見抜く力が重要であって、そのために数学やプログラミングからロジックを強化してはどうかと述べた。現状に対して批判のように聞こえた部分があったかもしれないが、そのような危機意識をもてば、さらなる発展につながるものと確信している。柴田里程先生が唱えた、変化に追従できるデータサイエンスが実現できるのではないかと。

### 参考文献

- [1] 今野浩, 鈴木久敏, 『整数計画法と組合せ最適化』, 日科技連, 1982.
- [2] J. Suzuki, “Learning bayesian belief networks based on the minimum description length principle: An efficient algorithm using the B & B technique,” *Machine Learning, Proceedings of the Thirteenth International Conference (ICML '96)*, Bari, Italy, July 3-6, 1996. Morgan Kaufmann, pp. 462-470, 1996.
- [3] 柴田里程, “データサイエンスのすすめ,” *日本統計学会誌*, **30**, pp. 327-332, 2000.
- [4] データサイエンス協会, データサイエンス スキルチェックリスト ver.3, [https://www.datascientist.or.jp/common/docs/skillcheck\\_ver3.00.pdf](https://www.datascientist.or.jp/common/docs/skillcheck_ver3.00.pdf) (2020年6月30日閲覧)
- [5] 鈴木讓, 『機械学習の数理 100問シリーズ 1 統計的機械学習の数理 100問 with R』, 共立出版, 2020.
- [6] 鈴木讓, 『機械学習の数理 100問シリーズ 2 統計的機械学習の数理 100問 with Python』, 共立出版, 2020.
- [7] 鈴木讓, 『機械学習の数理 100問シリーズ 3 スパース推定 100問 with R』, 共立出版, 2020. (出版予定)
- [8] 鈴木讓, 『機械学習の数理 100問シリーズ 4 スパース推定 100問 with Python』, 共立出版, 2020. (出版予定)
- [9] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [10] N. Johnson, “A dynamic programming algorithm for the fused Lasso and L 0-Segmentation,” *Journal of Computational and Graphical Statistics*, **22**, pp. 246-260, 2013.
- [11] R. Tibshirani and J. Taylor, “The solution path of the generalized lasso,” *Annals of Statistics*, **39**, pp. 1335-1371, 2011.
- [12] S. Boyd, N. Parikh, Eric Chu, B. Peleato and J. Eckstein, “Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers,” in *Foundations and Trends in Machine Learning*, **3**(1), pp. 1-122, 2010.