単層パーセプトロンの許容パラメータ領域の計算アルゴリズム

篠原 一成*・菊地 俊一郎**

大石 泰章**・木村 英紀*

An Algorithm to Compute the Admissible Parameter Region of a Single-Layered Perceptron

Issei SHINOHARA*, Syunichirou KIKUCHI**,
Yasuaki OISHI** and Hidenori KIMURA*

This paper provides an algorithm to compute the admissible parameter region in the learning by a single-layered perceptron. This algorithm is carried out by solving linear programming problems, the number of which is proportional to that of given samples.

Key Words: single-layered perceptron, Bayesian statistics, statistical mechanics, admissible region, linear programming

1. はじめに

近年学習機械の理論において, Bayes 統計の枠組を利用し, 例題が与えられたときの学習の成果を学習機械のパラメータの事後分布の形でとらえる統計力学的アプローチが注目されている^{3), 4)}. しかし, 一般の場合にパラメータの事後分布を求める計算は, 複雑な積分を伴い解析的にも数値的にも困難なことが多い.

上記のアプローチにおいて, 事後分布は Gibbs 分布を用いて構成され, 統計力学では温度の逆数を意味するパラメータ β が学習機械の例題への敏感さを規定する. たとえば $\beta = 0$ の場合には例題の持つ情報は全く反映されず, 事後分布は事前分布と同じである. 一方, $\beta = \infty$ の場合には, 事後分布は例題と合致する入出力関係を与えるパラメータの領域においてのみ非零の値をとる. この領域を許容領域と呼ぶ.

文献 1), 2) では, 線形関数による二分割を行なう単層パーセプトロンについて, 許容領域の性質を調べている. それによると, この場合の許容領域は複数の超平面で囲まれた多角錐の領域となるが, 許容領域を囲む超平面の数の平均は例題数に依存しない定数になる. このことから, 上記の場合には許容領域を効率良くかつ厳密に求められるのではないかと示唆を得る. 事後分布は許容領域の外では零であり, 許容領域の中では事前分布に比例するため, 許容領域さえ求めれば直ちに事後分布を求められることに注意されたい.

本論文では以上の考えにもとづき, 単層パーセプトロンの許容領域を, 線形計画法を使って効率よく求めるアルゴリズム

を提案し, 数値実験によってその有効性を実証する.

2. 問題設定

本論文で扱う学習機械は, 入力 $\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbf{R}^n$ と出力 $y \in \{-1, 1\}$ を持ち, その入出力関係はパラメータ $\mathbf{w} = (w_1, \dots, w_n)^T \in \mathbf{R}^n$ を用いて

$$y = \text{sign}(\mathbf{w} \cdot \mathbf{x}) \quad (1)$$

と表されるものとする. ただし,

$$\text{sign}(u) = \begin{cases} 1 & (u \geq 0), \\ -1 & (u < 0) \end{cases} \quad (2)$$

とする.

l 個の例題 $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)$ を与えたとき, 許容領域を

$$\{\mathbf{w} | y_1(\mathbf{w} \cdot \mathbf{x}_1) > 0, \dots, y_l(\mathbf{w} \cdot \mathbf{x}_l) > 0\} \quad (3)$$

と定める. 許容領域に属す $\mathbf{w}$ が, 例題に合致した入出力関係を与えることは明らかである. (3) 式中の l 個の不等式はそれぞれパラメータ空間 $\mathbf{R}^n$ 中の超平面を境界とし, 許容領域はこれらの超平面に囲まれた, 原点を頂点とする多角錐になる. 与えられる例題によっては, 許容領域が空集合になることもありうる.

このとき, l 個の超平面がすべて許容領域を表現するのに必要とは限らない. すなわち l 個の超平面の中には, それが存在しなくても許容領域が変化しないものもあり, このような超平面に対応する例題を無効例題と呼ぶ. そしてそれ以外の, 許容領域を表現するのに必要な例題を有効例題と呼ぶ. 与えられた例題に対して有効例題がすべて求めれば, 許容領域に関する無駄のない表現が得られたことになる. 以下で説明するのは, 線形計画法を用いて有効例題を求めるアルゴリズムである.

3. アルゴリズム

本アルゴリズムでは, 例題の数を一つずつ増やしそのたびごとに有効例題の集合を更新する方法をとる. いま, j 個の例題 $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_j, y_j)$ を与えたときの有効

* 東京大学大学院新領域創成科学研究科 東京都文京区本郷 7-3-1

** 東京大学大学院工学系研究科 東京都文京区本郷 7-3-1

* Graduate School of Frontier Sciences, University of Tokyo, Bunkyo-ku, Tokyo

** Graduate School of Engineering, University of Tokyo, Bunkyo-ku, Tokyo

(Received January 31, 2000)

(Revised May 1, 2000)

例題が $(x_{m_1}, y_{m_1}), \dots, (x_{m_k}, y_{m_k})$ の k 個であったとする ($k \leq j$). ここに $j+1$ 個目の例題 (x_{j+1}, y_{j+1}) を加えると, 許容領域は $\{w | y_{m_1}(w \cdot x_{m_1}) > 0, \dots, y_{m_k}(w \cdot x_{m_k}) > 0, y_{j+1}(w \cdot x_{j+1}) > 0\}$ となる. これに対する有効例題を求めるには, 以下の二点を調べる必要がある:

- (i) 許容領域が空集合にならないかどうか;
- (ii) $k+1$ 個の例題の中に無効例題が含まれていないかどうか.

上記の (i) については, 次のような線形計画問題を解けば調べるができる:

$$\begin{aligned} \min. \quad & z = w_1; \\ \text{s.t.} \quad & y_i(w \cdot x_i) \geq 1, \quad i = m_1, \dots, m_k, j+1. \end{aligned} \quad (4)$$

[命題] 許容領域が空集合であることの必要十分条件は, 線形計画問題 (4) が許容解を持たないことである. $\square$

目的関数は仮に $z = w_1$ としたが, この関数自体に意味はなく, 何を用いてもよい. いずれにせよ, 許容解の有無を調べれば, 許容領域が空集合かどうか分かる. 許容領域が空集合ならば, 学習は不可能ということを意味するので, その時点で学習は終わりである.

許容領域が空集合でないならば, 上記の (ii) を調べる. まず最初に, (x_{j+1}, y_{j+1}) が有効例題かどうかを調べる. 次のような線形計画問題を解くことで, これを調べることができる.

$$\begin{aligned} \min. \quad & z_{j+1} = y_{j+1}(w \cdot x_{j+1}); \\ \text{s.t.} \quad & y_i(w \cdot x_i) \geq 0, \quad i = m_1, \dots, m_k. \end{aligned} \quad (5)$$

[命題] 例題 (x_{j+1}, y_{j+1}) が無効なとき, 線形計画問題 (5) の最適解は $z_{j+1} = 0$ である. 例題 (x_{j+1}, y_{j+1}) が有効なとき, (5) の最適解は負の無限大である.

(証明) まず, (x_{j+1}, y_{j+1}) が無効例題のときは, 制約式で定まる領域内の点はすべて $y_{j+1}(w \cdot x_{j+1}) \geq 0$ を満たし, 特に原点 $w = 0$ のみ等号が成立する. よって目的関数 z_{j+1} の最小値は 0 となる. 一方, (x_{j+1}, y_{j+1}) が有効例題のとき, 制約式で定まる領域は $y_{j+1}(w \cdot x_{j+1}) \geq 0$ を満たす領域と, $y_{j+1}(w \cdot x_{j+1}) < 0$ を満たす領域とに二分される. 制約式で定まる領域内の点 w に対し, αw ($\alpha > 0$) もその領域に含まれるので, 目的関数の最小値は負の無限大である. $\square$

(x_{j+1}, y_{j+1}) が無効例題であれば, 許容領域は変化しないので, k 個の不等式を保持したまま次の例題 (x_{j+2}, y_{j+2}) の処理に移る. (x_{j+1}, y_{j+1}) が有効例題であれば, 次に $(x_{m_1}, y_{m_1}), \dots, (x_{m_k}, y_{m_k})$ という k 個の例題が, それぞれ有効例題かどうか調べる. これは例題 (x_{j+1}, y_{j+1}) の場合と同様に線形計画法を使って行なうことができる.

以上をまとめると, 次のようなアルゴリズムが得られる.

[アルゴリズム]

入力: l 個の例題

出力: 有効例題の集合 S

- (i) $S := \emptyset, j := 2$.
- (ii) 例題 (x_j, y_j) を加え, これによって許容領域が空集合にならないかどうかを調べる. 空集合であればアルゴリズムを終了する.
- (iii) 例題 (x_j, y_j) が有効かどうかを調べる. 有効ならステップ (iv), 無効ならステップ (vi) へ.
- (iv) それまでの有効例題 (S の要素) が有効かどうか調べる.
- (v) S に例題 (x_j, y_j) を加え, ステップ (iv) で無効とされた例題を削除する.
- (vi) $j = l$ ならば終了する. そうでなければ $j := j+1$ としてステップ (ii) に戻る. $\square$

上記のアルゴリズムの計算量について考察する. 例題 $x_1, \dots, x_l$ は $\|x_i\| = 1$ と規格化されているものとし, 半径 1 の $n-1$ 次元球面上の一様分布に従う独立な確率変数であるとする. また真のパラメータ w_* が存在し, 各 y_i は $\text{sign}(w_* \cdot x_i)$ で与えられるものとする. このとき文献 1) の定理 3.6, 系 3.7, または文献 2) の定理 3, 系 1 より次が成り立つ.

[命題] 上記の場合, 有効例題の数の期待値は十分大きな l に対して

$$\left[\frac{n}{\pi^n} \int_0^{\frac{\pi}{2}} \sin^{n-1} \phi d\phi \right]^{-1} \quad (6)$$

以下である. $\square$

これより上記のアルゴリズムにおいて例題を一つ増やして有効例題を更新するときに解かなくてはならない線形計画問題の数は平均的に定数以下である. したがって入力例題数が l のとき, アルゴリズム全体で解くべき線形計画問題の数は平均的に $O(l)$ である.

4. 数値実験

本節では上記アルゴリズムを適用した数値実験の結果を示す.

実験方法

1. 例題 (x, y) のうち, 入力 $x = (x_1, \dots, x_n)^T$ は, 半径 1 の $n-1$ 次元球面上の一様分布に従う独立な確率変数であるとする.
2. 例題の出力 y は, $w_* = (1, \dots, 1, 1)^T$ を用いて $y = \text{sign}(w_* \cdot x)$ により定める.
3. 入力空間の次元 $n = 4, 5, 6, 7, 8, 9$ の場合について, 例題数を 2000, 4000, 6000, 8000, 10000 個としたときの有効例題数を上記アルゴリズムを用いて計算する. また, そのときに解いた線形計画問題の数をカウントする.

実験結果

Fig. 1 は入力空間が 5, 9 次元の場合の例題数と有効例題数の関係を示したものである. ここで, エラーバーは実験を

Fig. 1 Number of active samples when the input dimension is five and nine

Fig. 2 Number of solved linear programming problems when the input dimension is five and nine

Fig. 3 Relation between the input dimension and the number of active samples (among 10000 total samples)

10 回繰り返したときの標準偏差を表す。この図から、上記アルゴリズムを用いたとき、例題数が十分多くなると、実際に有効例題数が定数に落ち着くことがわかる。また、そのときに線形計画問題を何回解いたかを Fig. 2 に示した。この図より、入力例題数が l のとき、アルゴリズム全体で解くべき線形計画問題の数は平均的に $O(l)$ であることも示せた。Fig. 3 には、10000 個の例題を与えたときの有効例題数の次元依存性を示した。有効例題数は次元ごとに定まる定数に落ち着くが、その定数は実験的には次元に比例することがわかる。

なお、入力 $\mathbf{x} = (x_1, \dots, x_n)^T$ の各成分がそれぞれ独立に区間 $[-2, 2]$ の一様分布に従う場合についても、同様の実験を行なった。この場合、前節末尾の命題は適用できないが、上と同様の実験結果が得られ、例題数が多くなると有効例題数は定数に落ち着くこと、解くべき線形計画問題の数は平均的に $O(l)$ であること、有効例題数は次元に比例することが確認できた。

5. おわりに

本論文では、統計力学的アプローチにおいて $\beta = \infty$ (温度零) の極限を考え、事後分布が非零の値をとる集合である許容領域を求めるアルゴリズムを与えた。求めた許容領域からは、Gibbs アルゴリズムや Bayes アルゴリズムなどの方法を用いて、テスト入力に対する出力の推定を行なうこともできる。計算量に関しては、本アルゴリズムは例題数 l に比例する回数だけ線形計画問題を解くことで実行できる。このことは数値実験によっても確認した。本アルゴリズムは線形関数に基づく単層パーセプトロンの場合にしか適用できない。しかし近年研究が盛んなサポートベクターマシーン⁵⁾においては入出力に非線形の変換を施すことによって上記のような単層パーセプトロンを非線形の学習に用いて成果をあげている。有効例題の概念はサポートベクターと関係が深く、同様の手法によって本論文の結果を非線形の学習に拡張できるかどうかは今後の検討課題である。

参考文献

- 1) 池田 和司：学習機械の予測誤差，東京大学大学院工学系研究科計数工学専攻学位論文 (1994)
- 2) K. Ikeda and S. Amari: Geometry of Admissible Parameter Region in Neural Learning, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Science, **E79-A-6**, 938/943 (1993)
- 3) E. Levin, N. Tishby and S. A. Solla: A Statistical Approach to Learning and Generalization in Layered Neural Networks, Proceedings of the 2nd Workshop on Computational Learning Theory, 245/260 (1989)
- 4) M. Oppor and D. Haussler: Calculation of the Learning Curve of Bayes Optimal Classification Algorithm for Learning a Perceptron with Noise, Proceedings of the 4th Workshop on Computational Learning Theory, 75/87 (1991)
- 5) V. N. Vapnik: Statistical Learning Theory, John Wiley & Sons (1998)