ロバスト最適化から見た機械学習

武田　朗子

機械学習の分野では，データから規則性やパターンを発見するため，しばしば数理最適化手法が用いられている．本稿では，われわれの成果も含めた 2 つの研究を取り上げ，ロバスト最適化がどのように機械学習において使われたかを紹介する．この研究成果の紹介を通して，「数理最適化の研究者（私）が機械学習分野で何ができたか」をお伝えしたい．

キーワード：ロバスト最適化，機械学習，2 値判別問題，サポートベクターマシン，ミニマックス確率マシン

1. はじめに

機械学習 (machine learning) は「見えている情報（データ）を手がかりに，見えていないのものを予測する技術」と言われている．典型的には，数字，文字，画像などのデータから，規則性・パターンを発見し，現状を把握や将来の予測をする．データから規則性やパターンを発見する過程は，しばしば数理最適化問題として定式化され，最適化手法を用いて解かれている．

機械学習の研究会には，統計，計算機科学，統計物理，アルゴリズム，最適化，さまざまなバックグラウンドを持った研究者が集まって，活発に交流している．私は最初に機械学習関連の研究会に顔を出し，研究成果を聞いたとき，「ここでこんなに数理最適化手法が使われているのか」と驚いた．それ以来，オペレーションズ・リサーチ，とりわけ数理最適化の専門家として，機械学習分野で何か面白い研究ができないものかと目論んでいる．機械学習分野には多くの若い研究者達が活発に研究を行っている．最近の数理最適化の関連研究をよく勉強し，どんどん研究に取り入れており，研究スピードがとても速く感じる．少し前には，機械学習分野において，二次錐計画法，半正定値計画法という言葉をよく聞いた．最近は，劣モジュラ最適化，DC (difference of convex functions) 最適化を用いた研究をよく目にする．

ここ最近，機械学習分野では，「ロバスト最適化法」を用いた研究成果がいくつか発表されている．本稿では，Xu-Caramanis-Mannor の成果 [13]，そしてわれわれの成果 [11] について紹介したい．ちなみに [13] の著者の一人である Caramanis も私も，最適化分野でロバスト最適化法を研究していたというバックグラウンドを持っている．機械学習分野において応用問題を持っているわけでもなく，特別なデータを持っているわけではない私にとって，新しい機械学習モデルを考案するには少々敷居が高く，また流行りのモデルに対して効率的なアルゴリズムを考案しようにも機械学習研究者のスピードについていけずに乗り遅れている状況である．そこで，私にとって参入しやすく，また，機械学習分野に多少なりとも貢献できる研究としては，“機械学習分野においてまだそれほど知られていない最適化のツールを使って，よく知られた機械学習モデルの性質をより深く調べる”ような研究，つまり既存モデルをより深く掘り下げる研究であった．

このような研究スタイルは賛否両論のコメントをもらいやすい．かつて，既存の 2 値判別モデルの予測性能を理論的に評価した研究論文 [12] を投稿した際に，ある査読者からは「今まで数値実験を通してモデルの良さは示されていたが理論的にモデルの良さを示した初めての論文だ」と手放しの賛辞を送られ，別の査読者からは「この研究は新しいモデルも新しいアルゴリズムも提案していない」との批判をいただいた．機械学習分野において，王道の研究スタイルではないかもしれないが，既存モデルをより深く掘り下げる研究も大切だと考えて研究を行っている．

本稿では，2 値判別問題の紹介，既存手法の解説をするとともに，機械学習分野においてロバスト最適化法を用いた研究（Xu らの成果 [13] とわれわれの成果 [11]）を紹介することで，「数理最適化の研究者が機械学習分野で何ができたか」をお伝えしたい．

たけだ　あきこ
東京大学大学院 情報理工学系研究科
〒 113–8656 東京都文京区本郷 7–3–1

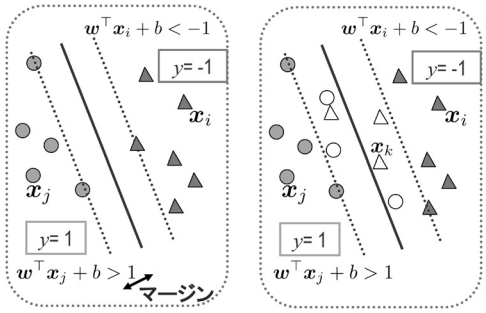

図 1 線形分離可能なデータ集合（左）と線形分離不可能なデータ集合（右）

2. 2 値判別問題

2 値判別問題とは，複数のデータが二つのグループに分かれている状態で新たな未知データが与えられたときに，そのデータがどちらのグループに属しているかを決める問題である．ベクトルとラベルの組：$(\boldsymbol{x}_i, y_i)$, $i \in M := \{1, \ldots, m\}$ が与えられており，y_i は -1 または 1 の 2 値をとるラベルで，$\boldsymbol{x}_i$ は i 番目のデータベクトルを表すものとする．“学習”とは，これらのデータに何らかの基準で最も合う関数 $y = h(\boldsymbol{x})$ を求めることである．この関数を用いて，未知のデータ $\hat{\boldsymbol{x}}$ のラベルを $y = h(\hat{\boldsymbol{x}})$ と予測できる．ここでは簡略化のために，線形関数に基づく判別関数 $h(\boldsymbol{x}) = \mathrm{sign}(\boldsymbol{w}^\top \boldsymbol{x} + b)$ に限定して話を進めたい．ここで，$\boldsymbol{w} \in \mathbb{R}^n$，$b \in \mathbb{R}$，そして $\mathrm{sign}(\xi)$ は $\xi \geq 0$ のときに 1，$\xi < 0$ では -1 をとるものとする．

与えられたデータに合わせすぎた予測関数 $h(\boldsymbol{x})$ を得てしまうことを“過学習”と呼ぶ．汎化誤差（学習に使わなかった未知のデータに対する予測誤差）をいかに小さくするかが機械学習の課題である．機械学習モデルの自由度に抑制を加えて過学習を防ぐため，正則化項（例えば，$\|\boldsymbol{w}\|^2$）を含んだ定式化がなされることが多い．

3. 代表的な判別手法

さまざまな判別手法が提案されているが，ここでは後の議論に必要な判別手法のみを挙げておく．

3.1 サポートベクターマシン (SVM)

サポートベクターマシン (SVM) は現在知られている多くの手法の中でも最も判別性能の優れた学習手法の一つである．

図 1（左図）に示すように，▲のグループと●のグループに分離可能なデータ集合が与えられていた場合，SVM では，分離超平面（ここでは直線）とデータ間の距離：

$$f(\boldsymbol{w}, b; \boldsymbol{x}, y) := \frac{y(\boldsymbol{w}^\top \boldsymbol{x} + b)}{\|\boldsymbol{w}\|}$$

を用いて，すべてのデータに対する最小値 ($\min_{i \in M} f(\boldsymbol{w}, b; \boldsymbol{x}_i, y_i)$，これをマージンと呼ぶ) が $(\boldsymbol{w}, b)$ について最大になるように分離超平面が求められる．これを定式化すると以下の問題になる：

$$\max_{\boldsymbol{w}, b} \min_{i \in M} f(\boldsymbol{w}, b; \boldsymbol{x}_i, y_i)$$

また，これを変形して

$$\min_{\boldsymbol{w}, b} \frac{1}{2}\|\boldsymbol{w}\|^2 \quad \text{s.t. } y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) \geq 1, \ \ i \in M \ (1)$$

という凸 2 次計画問題に帰着される．(1) はハードマージン (hard margin) SVM [3] と呼ばれる．マージンを最大にするような分離超平面が最も汎化能力の高い（つまり汎化誤差を最小にするような）超平面であることが知られている．

実問題で線形分離可能な場合は稀であり，(1) の制約を緩める工夫が必要である．そのような代表的なモデルとして，C-SVM [4]:

$$\min_{\boldsymbol{w}, b, \boldsymbol{\xi}} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{m} \xi_i$$
$$\text{s.t. } y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \ \ i \in M$$

や ν-SVM [10]:

$$\min_{\boldsymbol{w}, b, \boldsymbol{\xi}, \rho} \frac{1}{2}\|\boldsymbol{w}\|^2 - \nu\rho + \frac{1}{m}\sum_{i=1}^{m} \xi_i$$
$$\text{s.t. } y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) \geq \rho - \xi_i, \quad \xi_i \geq 0, \ \ i \in M$$

が知られている．正の値をもつ ξ_i は，線形分離の違反に対応しており，損失とみなされるものである．図 1（右図）の△や○（例えば，$\boldsymbol{x}_k$）に対応する ξ_k が正の値をとる．C-SVM と ν-SVM は，マージン最大化と経験損失最小化の二つの目的をコントロールするために，それぞれ，C (> 0) と ν $(\in [0, 1])$ という正値パラメータを含んだ定式化がなされている．

C-SVM と ν-SVM は凸 2 次計画問題であり，SMO (Sequential Minimal Optimization) といった効率的な解法が提案されている．パラメータ変換により，ν-SVM と C-SVM は基本的に同じモデルとなることが示されている [10].

線形分離不可能なデータセットに対して，ν-SVM の

パラメータ ν をある下限値 $\nu_{\min}$ 以下に設定すると，最適解 $(\boldsymbol{w}, b)$ がすべて 0 になってしまう（詳細は [5] 等を参照のこと）．$\nu = 0$ まで取れるように ν-SVM を拡張したモデル (Eν-SVM [9]) も提案されている．

3.2 ミニマックス確率マシン (MPM)

次に，Lanckriet ら [7] によって提案されたミニマックス確率マシン (Minimax Probability Machine; MPM) を紹介する．MPM では，2 値判別問題の各クラスの入力データとして，n 次元の確率変数 $\boldsymbol{x}_+$，$\boldsymbol{x}_-$ が用いられ，また，それぞれについて平均 $\bar{\boldsymbol{x}}_+$，$\bar{\boldsymbol{x}}_- \in \mathbb{R}^n$ と分散共分散行列 Σ_+，$\Sigma_- \in \mathbb{R}^{n\times n}$ が与えられているものとする．ここで，分散共分散行列は正定値対称行列と仮定する．

この与えられた平均と分散共分散行列をもつあらゆる分布に対して，最も高い確率で二つのクラスのデータを分けるように超平面 $\boldsymbol{w}^\top \boldsymbol{x} + b = 0$ を決定することが目的である．これを定式化すると以下の問題となる．

$$\begin{aligned} \max_{\alpha, \boldsymbol{w}, b} \quad & \alpha \\ \text{s.t.} \quad & \min_{\boldsymbol{x}_+ \sim (\bar{\boldsymbol{x}}_+, \Sigma_+)} \Pr\{\boldsymbol{w}^\top \boldsymbol{x}_+ + b \geq 0\} \geq \alpha \\ & \min_{\boldsymbol{x}_- \sim (\bar{\boldsymbol{x}}_-, \Sigma_-)} \Pr\{\boldsymbol{w}^\top \boldsymbol{x}_- + b \leq 0\} \geq \alpha \end{aligned} \tag{2}$$

$\boldsymbol{x}_+ \sim (\bar{\boldsymbol{x}}_+, \Sigma_+)$ は，平均 $\bar{\boldsymbol{x}}_+$ と分散共分散行列 Σ_+ をもつある分布に確率変数 $\boldsymbol{x}_+$ が従うことを示す．$\boldsymbol{x}_-$ においても同様である．(2) は，判別に関して最悪 (min) な分布を想定した場合を最も良く (max) 判別することを表す．この定式化はミニマックス確率マシン (MPM) と呼ばれている．(2) は二次錐計画問題に変形できる．

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \|\Sigma_+^{1/2} \boldsymbol{w}\| + \|\Sigma_-^{1/2} \boldsymbol{w}\| \\ \text{s.t.} \quad & \boldsymbol{w}^\top (\bar{\boldsymbol{x}}_+ - \bar{\boldsymbol{x}}_-) = 1 \end{aligned} \tag{3}$$

ただし，$\Sigma_+^{1/2}$ は Σ_+ の平方根行列とする．さらに，(3) の最適解から (2) の最適解 (b^*, α^*) が求まる．

Nath-Bhattacharyya [8] は，マージン最大化の考え方を MPM に取り入れたモデルを提案した．判別誤りに対して許容する率を $\eta \in [0, 1]$ として，以下のように定式化される．

$$\begin{aligned} \min_{\boldsymbol{w}, b} \quad & \tfrac{1}{2}\|\boldsymbol{w}\|^2 \\ \text{s.t.} \quad & \max_{\boldsymbol{x}_+ \sim (\bar{\boldsymbol{x}}_+, \Sigma_+)} \Pr\{\boldsymbol{x}_+^\top \boldsymbol{w} + b < 0\} \leq \eta, \\ & \max_{\boldsymbol{x}_- \sim (\bar{\boldsymbol{x}}_-, \Sigma_-)} \Pr\{\boldsymbol{x}_-^\top \boldsymbol{w} + b > 0\} \leq \eta. \end{aligned} \tag{4}$$

このモデルをここでは MM-MPM と呼ぶことにする．MPM と同様に，MM-MPM もまた二次錐計画問題に変形できる．

4. ロバスト最適化

ここでは，ロバスト最適化について簡単に説明をし，ロバスト最適化の観点から正則化項に新しい解釈を与えた Xu らの成果 [13] について，簡単に紹介したい．

4.1 ロバスト最適化とは

現実の問題にはさまざまな不確実性が存在しており，現実の問題を数理最適化問題として定式化する際には，“測定誤差が含まれているデータ”や“将来の需要の代わりに過去のデータを用いた予測値”などを使わなければならないこともある．そこで，微小なデータの変動に対して強健な解を得ることを目的としたロバスト最適化法 [1] が，近年注目を集めている。ロバスト最適化では，不確実なデータの生じ得る範囲をあらかじめ設定し，その中で最悪の状況が生じた場合を想定したモデル化が行われている．ロバスト最適化による解は，不確実なデータが想定範囲内で動く分には制約式を破ることもなく目的関数値もひどく悪くなることはないため，微小な変動に対して強健な解を得ることができる．

ここでは，目的関数にのみ不確実なデータが含まれた意思決定問題として，以下の最適化問題を考える．

$$\min_{\boldsymbol{w} \in W} \quad f(\boldsymbol{w}, \widehat{\boldsymbol{x}}) \tag{5}$$

ここで，$\widehat{\boldsymbol{x}}$ は不確実なデータ，$\boldsymbol{w}$ は意思決定変数，$f(\boldsymbol{w}, \widehat{\boldsymbol{x}})$ は目的関数，W は実行可能領域とする．

問題 (5) の不確実なデータ $\widehat{\boldsymbol{x}}$ が生じうる範囲を不確実性集合と呼び，ここでは $\mathcal{U}$ と記述する．(5) に対するロバスト最適化問題は，次のように定式化される．

$$\min_{\boldsymbol{w} \in W} \quad \max_{\boldsymbol{x} \in \mathcal{U}} f(\boldsymbol{w}, \boldsymbol{x}) \tag{6}$$

不確実性集合 $\mathcal{U}$ の要素が無限にある場合には，問題 (6) において無限本の目的関数 $f(\boldsymbol{w}, \boldsymbol{x})$，$\forall \boldsymbol{x} \in \mathcal{U}$，を考慮することになる．(6) は，そのような目的関数の中から最悪状況を想定して，最もよい解を見つける問題である．

たとえ $\boldsymbol{w}$ の実行可能領域 W が凸集合で与えられても，$\mathcal{U}$ として一般的な集合を想定した場合には，ロバスト最適化問題を解きやすい最適化問題に帰着させることは難しい．しかし，矩形や楕円形などの扱いやすい不確実性集合 $\mathcal{U}$ を仮定すれば，(6) は解きやすい凸計画問題に帰着されることが知られている [1].

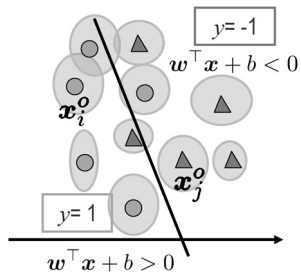

図 2　各データに不確実性集合を想定

4.2　正則化項とロバスト化の同値性

過学習を防ぐための工夫として，判別モデルの定式化に正則化項 $\|\boldsymbol{w}\|^2$ がしばしば用いられる．この正則化項をロバスト最適化の視点で解釈を与えたのが Xu ら [13] である．

Xu らは C-SVM の正則化項を除き，経験損失だけを最小化する問題

$$\min_{\boldsymbol{w},b} \sum_{i=1}^{m} [1 - y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b)]^+ \tag{7}$$

を扱っている（ただし，$[X]^+ := \max\{X, 0\}$）．

データ $\boldsymbol{x}_i (i = 1, \ldots, m)$ が誤差を含んでいて不確実であると仮定し，$\boldsymbol{x}_i$ の代わりに不確実性集合（所与のデータ $\boldsymbol{x}_i^0$，$i \in M$，を中心とした楕円の集合）：

$$\mathcal{U} = \left\{ (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) : \begin{array}{r} \boldsymbol{x}_i = \boldsymbol{x}_i^0 + \boldsymbol{\Delta x}_i,\ i \in M, \\ \sum_{i=1}^{m} \|\boldsymbol{\Delta x}_i\| \le \sigma \end{array} \right\}$$

を想定する．パラメータ σ により楕円の大きさが決められる．この経験損失最小化問題 (7) をロバスト化すると次の問題：

$$\begin{array}{ll} \min\limits_{\boldsymbol{w},b} \max\limits_{\boldsymbol{\Delta x}_i, i \in M} & \sum_{i=1}^{m} [1 - y_i\{\boldsymbol{w}^\top(\boldsymbol{x}_i^0 + \boldsymbol{\Delta x}_i) + b\}]^+ \\ \text{s.t.} & \sum_{i=1}^{m} \|\boldsymbol{\Delta x}_i\| \le \sigma \end{array}$$

が得られ，次の等価な問題に帰着される [13]．

$$\begin{array}{l} \min\limits_{\boldsymbol{w},b,\boldsymbol{\xi}} \sigma\|\boldsymbol{w}\| + \sum_{i=1}^{m} \xi_i \\ \text{s.t. } y_i(\boldsymbol{w}^\top \boldsymbol{x}_i^0 + b) \ge 1 - \xi_i, \quad \xi_i \ge 0, \quad i \in M \end{array} \tag{8}$$

問題 (8) の目的関数は“正則化項+損失項”から成り立っており，C-SVM と非常に似た問題となっている．経験損失最小化問題をロバスト化することによって正則化項が得られており，“ロバスト化＝正則化”という関係が示唆される．つまり，所与のデータ $\boldsymbol{x}_i^0, i \in M$ に対して過学習を避けるために正則化を取り入れた判別ルールを，データの不確実性を考慮したロバスト解を用いて構築することができる．

5.　ロバスト最適化に基づく判別モデル

本節では，Xu ら [13] とは異なる方法で，ロバスト最適化法を機械学習に適用することを試みる．

5.1　ロバスト判別モデル

ロバスト判別モデルを定式化するうえで，$\boldsymbol{x}_+$ をクラス 1 に対するデータの代表点（例えば，平均ベクトル），$\boldsymbol{x}_-$ をクラス -1 に対するデータの代表点とみなすことにする．また，$\boldsymbol{x}_+$ と $\boldsymbol{x}_-$ の生じ得る範囲をそれぞれ $\mathcal{U}_+$，$\mathcal{U}_-$ と記述し，観測データ $(\boldsymbol{x}_i, y_i)$，$i \in M$，を用いて構築する．

ロバスト判別モデルを次のように定式化する．

$$\max_{\boldsymbol{w}:\|\boldsymbol{w}\|=1} \min_{\boldsymbol{x}_+ \in \mathcal{U}_+, \boldsymbol{x}_- \in \mathcal{U}_-} (\boldsymbol{x}_+ - \boldsymbol{x}_-)^\top \boldsymbol{w} \tag{9}$$

この最適解を $\boldsymbol{w}^*$ とし，(9) の内側の最小化問題の最適解 $\boldsymbol{x}_+^*$ と $\boldsymbol{x}_-^*$ を用いて b^* を適切に求める（詳細については [11] を参照）．不確実性集合 $\mathcal{U}_+$，$\mathcal{U}_-$ としてあるタイプの凸集合を想定すると，ロバスト判別モデルと既存モデルの判別関数 $h(\boldsymbol{x})$ が一致することを示すことができる（表 1 を参照）．

ロバスト判別モデル (9) は非凸な制約式 $\|\boldsymbol{w}\|^2 = 1$ を含んでおり，一見，求解が難しく見える．しかし，問題の難しさは $\mathcal{U}_+$ と $\mathcal{U}_-$ に交わりがあるか否かに依存する．もし，図 3（左図）が示すように，$\mathcal{U}_+$ と $\mathcal{U}_-$ に交わりがない場合には，凸制約式 $\|\boldsymbol{w}\|^2 \le 1$ に変えても最適解は変わらない．つまり，(9) の制約式を $\|\boldsymbol{w}\|^2 \le 1$ に変えて，凸最適化問題を解けばよい．しかし，図 3（右図）が示すように，$\mathcal{U}_+$ と $\mathcal{U}_-$ に交わりがある場合には，(9) の制約式を $\|\boldsymbol{w}\|^2 \ge 1$ に変えることができるものの，依然として非凸最適化問題のままである．この問題に対して局所最適解を求めるための解法 [9, 11] が提案されている．

5.2　既存の判別モデルとの関係

ロバスト判別モデルに必要な入力データである不確実性集合，つまり (9) の $\mathcal{U}_+$ と $\mathcal{U}_-$ の例を紹介する．$\mathcal{U}_+$，$\mathcal{U}_-$ として 2 種類の楕円体や凸多面体を採用した場合，表 1 が示すように，それぞれが既存の判別モデルと対応していることを簡単に述べる．

■**ハードマージン SVM**：　与えられたデータセット $(\boldsymbol{x}_i, y_i)$，$i \in M$ を $M_+ = \{i \in M : y_i = 1\}$ と

表 1 ロバスト判別モデルと既存の判別モデルとの関係（詳細は [11] を参照）．× はそのケースが生じないことを表し，√ は対応する既存モデルがないことを表す．

不確実性集合 $\mathcal{U}_\pm$	$\mathcal{U}_+$ と $\mathcal{U}_-$ の関係		
	交わらない	接する	真に交わる
楕円-a (12)	MM-MPM [8]	MPM [7]	√
楕円-b (14)	FS-FDA [2]	FDA [6]	√
縮退凸包 (11)	ν-SVM [10]	√	Eν-SVM [9]
凸包 (10)	ハードマージン SVM [3]	×	×

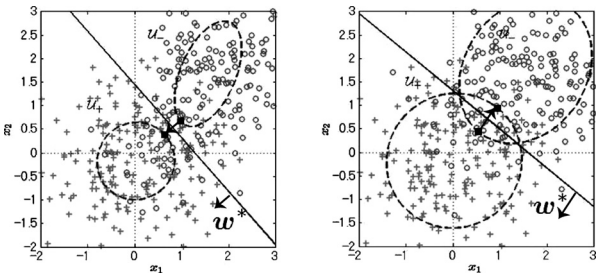

図 3 $\mathcal{U}_+$ と $\mathcal{U}_-$ が交わりを持たないケース（左図）と $\mathcal{U}_+$ と $\mathcal{U}_-$ が交わりを持つケース（右図）．直線は (9) の最適解による判別平面を示し，黒い四角は最適解 $\boldsymbol{x}_+^* \in \mathcal{U}_+, \boldsymbol{x}_-^* \in \mathcal{U}_-$ を示す．

$M_+ = \{i \in M : y_i = -1\}$ の二つのクラスに分け，それぞれのクラスに対する $\mathcal{U}_+$（$\mathcal{U}_-$ も同様）を以下のように構築する．

$$\left\{ \sum_{i \in M_+} \lambda_i \boldsymbol{x}_i \ : \ \sum_{i \in M_+} \lambda_i = 1,\ 0 \le \lambda_i, i \in M_+ \right\}. \quad (10)$$

データセットが線形分離可能，つまり，$\mathcal{U}_+ \cap \mathcal{U}_- = \emptyset$ のときには，ロバスト判別モデル (9) とハードマージン SVM (1) は一致する．

■ **ν-SVM と Eν-SVM**： また，縮退凸包 (reduced convex hulls) [5] を用いて，$\boldsymbol{x}_+$ の不確実性集合 $\mathcal{U}_+^\nu$ を以下のように定義する．

$$\left\{ \sum_{i \in M_+} \lambda_i \boldsymbol{x}_i : \sum_{i \in M_+} \lambda_i = 1,\ 0 \le \lambda_i \le \frac{2}{\nu m}, i \in M_+ \right\}. \quad (11)$$

$\boldsymbol{x}_-$ の不確実性集合 $\mathcal{U}_-^\nu$ も同様に定義する．

3.1 節で導入した，ν-SVM で取りうるパラメータ ν の下限値 $\nu_{\min}$ を用いると，$\mathcal{U}_+^{\nu_{\min}}$ と $\mathcal{U}_-^{\nu_{\min}}$ は接する縮退凸包となる．以下の命題が成り立つ．

命題 5.1 ([11])**.** パラメータ $\nu > \nu_{\min}$ を用いて作られた $\mathcal{U}_+^\nu$ と $\mathcal{U}_-^\nu$ は $\mathcal{U}_+^\nu \cap \mathcal{U}_-^\nu = \emptyset$ であり，ロバスト判別モデル (9) は ν-SVM と等価である．$\nu \le \nu_{\min}$ の場合には $\mathcal{U}_+^\nu$ と $\mathcal{U}_-^\nu$ は交わりを持ち，(9) は Eν-SVM と等価である．

■ **MPM と MM-MPM**： ここでは，不確実性集合として，それぞれ，中心を $\bar{\boldsymbol{x}}_\pm$ に持ち，正定値行列 $\Sigma_\pm$ で形が定まる楕円：

$$\mathcal{U}_+^\kappa = \{\bar{\boldsymbol{x}}_+ + \Sigma_+^{1/2}\boldsymbol{u} : \|\boldsymbol{u}\| \le \kappa\} \quad (12)$$

と同様に定義した $\mathcal{U}_-^\kappa$ を考える．この不確実性集合のもとで，(9) は次の問題に帰着される．

$$\min_{\boldsymbol{w}:\|\boldsymbol{w}\|^2=1} \kappa\|\Sigma_+^{\frac{1}{2}}\boldsymbol{w}\| + \kappa\|\Sigma_-^{\frac{1}{2}}\boldsymbol{w}\| - \boldsymbol{w}^\top(\bar{\boldsymbol{x}}_+ - \bar{\boldsymbol{x}}_-) \quad (13)$$

$\mathcal{U}_+^\kappa$ と $\mathcal{U}_-^\kappa$ が接するようなパラメータ κ の値を $\kappa_{\max}$ とする．問題 (13) は非凸計画問題でありこのままでは解くことが難しいようにみえるが，$\kappa < \kappa_{\max}$ の場合には二つの楕円は交わりを持たず，最適解を変えることなく非凸制約式 $\|\boldsymbol{w}\|^2 = 1$ を凸制約式 $\|\boldsymbol{w}\|^2 \le 1$ に置き換えることができる．

命題 5.2 ([11])**.** $\kappa \in [0, \kappa_{\max})$ の場合には，(13) は MM-MPM (4) と等価であり，$\kappa = \kappa_{\max}$ の場合には，(13) は MPM (3) と等価である．

■ **FDA と FS-FDA**： 誌面の都合上，割愛するが，表 1 の FDA [6] や FS-FDA（FDA に基づく特徴選択法）[2] は，不確実性集合

$$\mathcal{U}^\varsigma = \{\boldsymbol{x} = (\bar{\boldsymbol{x}}_+ - \bar{\boldsymbol{x}}_-) + (\Sigma_+ + \Sigma_-)^{1/2}\boldsymbol{u} : \|\boldsymbol{u}\| \le \varsigma\} \quad (14)$$

を用いたロバスト判別モデル (9) として，表すことができる。

6. おわりに

表 1 に示したように，ロバスト最適化による定式化 (9) を用いていくつかの既存の判別モデルをつなげることができた．入力データや定式化が全く異なる既存モデル（SVM や MPM）がロバスト最適化問題として記述でき，それらの違いは不確実なデータ $\boldsymbol{x}_+$ と $\boldsymbol{x}_-$ に対して想定する範囲（$\mathcal{U}_+$，$\mathcal{U}_-$）にある．これに気づいたときには，非常に面白い知見が得られたように感じた．数ある既存モデルの関係が明らかになり，さらに，うまく $\mathcal{U}_+$，$\mathcal{U}_-$ を設定すれば，よりよい判別モデルが得られる可能性もある．既存モデルの関係を探ることによるメリットがあると思われるが，研究スピードが早く，どんどん新しい数理モデルが生まれる分野ではこういった研究はなかなか評価されない．実際に，既存モデルを関係づけただけでは評価してもらえず，表 1 の $\surd$ に対応する新しい判別モデルを提案し，数値実験を通して「どのようなときにこの新しいモデルが有効か」を示すことで，ようやく評価してもらうことができた．

数理最適化の知識をウリにして機械学習分野で研究を行うことに，今なお難しさを感じる．その一方で，機械学習分野には数理最適化法の応用先がいろいろとある．また，機械学習分野に出入りすることで，どのような最適化法が望まれているのかもわかる．異分野で研究を行うことは苦労もあるが，得られるものも多い．本稿を通して，“異分野で研究することの面白さ（大変さだけではなく…）”を感じていただけたら幸いである．

参考文献

[1] A. Ben-Tal, L. El-Ghaoui and A. Nemirovski, *Robust Optimization*, Princeton University Press, Princeton, 2009.

[2] C. Bhattacharyya, “Second Order Cone Programming Formulations for Feature Selection,” *Journal of Machine Learning Research*, **5**, 1417–1433, 2004.

[3] B. E. Boser, I. M. Guyon and V. N. Vapnik, “A Training Algorithm for Optimal Margin Classifiers,” *COLT*, pp. 144–152, ACM Press, 1992.

[4] C. Cortes and V. Vapnik, “Support-vector Networks,” *Machine Learning*, **20**, 273–297, 1995.

[5] D. J. Crisp and C. J. C. Burges, “A Geometric Interpretation of ν-SVM Classifiers,” *NIPS 12*, pp. 244–250, MIT Press, 2000.

[6] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, Boston, 1990.

[7] G. R. G. Lanckriet, L. El Ghaoui, C. Bhattacharyya and M. I. Jordan, “A Robust Minimax Approach to Classification,” *Journal of Machine Learning Research*, **3**, 555–582, 2002.

[8] J. S. Nath and C. Bhattacharyya, “Maximum Margin Classifiers with Specified False Positive and False Negative Error Rates,” *SDM*, pp. 35–46, SIAM, 2007.

[9] F. Perez-Cruz, J. Weston, D. J. L. Hermann and B. Schölkopf, “Extension of the ν-SVM Range for Classification,” *Advances in Learning Theory: Methods, Models and Applications 190*, pp. 179–196, Amsterdam, IOS Press, 2003.

[10] B. Schölkopf, A. Smola, R. Williamson and P. Bartlett, “New Support Vector Algorithms,” *Neural Computation*, **12**, 1207–1245, 2000.

[11] A. Takeda, H. Mitsugi and T. Kanamori, “A Unified Classification Model Based on Robust Optimization,” *Neural Computation*, **25**, 759–804, 2013.

[12] A. Takeda and M. Sugiyama, “ν-support Vector Machine as Conditional Value-at-risk Minimization,” *ICML 2008*, 1056–1063, 2008.

[13] H. Xu, C. Caramanis and S. Mannor, “Robustness and Regularization of Support Vector Machines,” *Journal of Machine Leaning Research*, **10**, 1485–1510, 2009.