6R-2

複素サポートベクターマシン

篠田 北斗 † 服部 元信 †† 小林 正樹 ††
† 山梨大学大学院医学工学総合教育部 †† 山梨大学大学院医学工学総合研究部

1 はじめに

近年，機械学習法の中でも Support Vector Machine (SVM) が注目を集めている．SVM における主要な理論であるマージン最大化法とカーネル法は，高次元特徴空間における適切な (汎化性能の良い) 決定境界を自動で獲得する．SVM は，二次計画問題に落としこむことで，これらの学習を計算量的に非常に効率よく解くことができる．これらの総合的な性能の高さから，SVM はパターン認識などの分野を中心に，様々な応用がなされてきた．また，扱う情報の種類に応じて，SVM 自身の様々な改良も行われている．例えば，学習器の出力が連続値である SVM として Support Vector Regression (SVR) がある．このように，扱う問題や信号の性質に依存して，様々なタイプの SVM が使い分けられている．

本稿では，SVM で扱う情報として複素数に着目し，各変数を複素数値化した複素サポートベクターマシン (CVSVM: Complex-Valued SVM) を提案する．この手法では，複素ニューラルネットワークの一種である Multi-Valued Neuron(MVN)[1] をベースにし，マージン最大化法とカーネル法を取り入れたものである．CVSVM の持つ特徴には，1) 出力は多値状態を取る複素数であること，2) マージン最大化学習によって得た未知の事例に対する識別汎化性能を持つこと，3) カーネル関数を用いた非線形クラス分類ができること，などである．計算機実験では，いくつかのベンチマークで提案法の有効性を調査する．

2. Multi-Valued Neuron

MVN(Multi-Valued Neuron)[1] は，複素数における多閾値を考慮した活性化関数を用いたニューラルネットワークである．この手法の出力は，次式で表される．

$$
\begin{aligned}
f(x_1,\cdots,x_n) =& P(w_0 + w_1x_1 + \cdots + w_nx_n)\\
=& P(\boldsymbol{w}^T\boldsymbol{x}) \qquad (1)
\end{aligned}
$$

ここで，$\boldsymbol{x} = (x_0, x_1, \cdots, x_n)^T$ は入力ベクトル ($x_0 = 1$ で固定)，$\boldsymbol{w} = (w_0, w_1, \ldots, w_n)^T$ は重みベクトル，

Complex-Valued Support Vector Machines
† Hokuto SHINODA
†† Motonobu HATTORI
†† Masaki KOBAYASHI
Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi (†)
Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi (††)

P は以下に示すニューロンの活性化関数である．

$$
\begin{aligned}
&P(z) = \exp(i2\pi j/K),\\
&\quad if \;\; 2\pi(j-0.5)/K \le \arg[z] < 2\pi(j+0.5)/K
\end{aligned} \qquad (2)
$$

ここで，$i = \sqrt{-1}$，変数 j は $j = 0, 1, \ldots, K-1$ と変化する．この活性化関数は，入力の複素数 z の偏角 $\arg[z]$ に応じて，対応する量子化点 (単位円を K 個に分割した交点) のいずれかに量子化する．

MVN には一般的な誤差訂正法による学習が適用できる．また，MVN の階層型ネットワークとして MLMVN(Multi-Layer MVN)[1] が提案されている．

3. Complex-Valued SVM

3.1. マージン最大化

マージンとは，分類器の決定境界に最も近い教師事例が，決定境界に対して持つ幾何学的な距離のことである．マージンをできる限り最大化することによって，教師事例から決定境界を遠ざけ，教師の周辺に安全域を確保する．これによって，未知のパターンへの汎化が期待できる．

提案法の CVSVM ではマージンを次式で表現する．

$$
\begin{aligned}
&\gamma = \min_{k,\theta} \frac{|\psi\left(s(\boldsymbol{x}_k),\theta\right)|}{\sqrt{\langle \boldsymbol{w}\cdot\boldsymbol{w}\rangle}},\\
&\qquad k = 1,\ldots,m,\; \theta = y_k\epsilon, y_k\epsilon^*.
\end{aligned} \qquad (3)
$$

ここで，$\boldsymbol{x}_k$ は m 個中 k 番目の教師事例，$s(\boldsymbol{x}_k) = \boldsymbol{w}^T\boldsymbol{x}_k$ はモデルの結合総和，$\epsilon = \pi/K$ は分割幅，$\psi(z,\theta) = \mathrm{Im}[z\theta^*]$ である．また，* は複素共役を示し，$\langle \boldsymbol{z}\cdot\boldsymbol{z}'\rangle = \sum_{j=0}^{n} z_j z_j'^*$ はエルミート内積を示す．操作 ψ を用いると，k 番目の教師事例 $\boldsymbol{x}_k$ が正しく分類される領域が次式によって定義できる．

$$
\begin{cases}
\psi(s(\boldsymbol{x}), y_k\epsilon^*) \ge 0\\
\psi(s(\boldsymbol{x}), y_k\epsilon) < 0
\end{cases}, \quad k = 1,\ldots,m \qquad (4)
$$

制約条件 (4) のもと，式 (3) を最大化すれば良いのだが，ここで問題に対して以下の条件を加える．

$$
\begin{aligned}
&\min_{k,\theta} |\psi\left(s(\boldsymbol{x}_k),\theta\right)| = 1,\\
&\qquad k = 1,\ldots,m,\; \theta = y_k\epsilon, y_k\epsilon^*
\end{aligned} \qquad (5)
$$

これは，重みのスケールを固定する役割を担い，問題の解に一意性を与える．式 (5) による制約条件を加味

すると，最適化問題を次式で定義できる．

$$\begin{aligned} \min_{\boldsymbol{w},\xi_k^-,\xi_k^+} \quad & \langle \boldsymbol{w}\cdot\boldsymbol{w}\rangle + C\sum_{k=1}^{m}\left(\xi_k^- + \xi_k^+\right) \\ \text{subject to :} \quad & \mathrm{Im}\left[(y_k\epsilon^*)^*\boldsymbol{w}^T\boldsymbol{x}_k\right] \geq 1-\xi_k^-, \\ & \mathrm{Im}\left[(y_k\epsilon)^*\boldsymbol{w}^T\boldsymbol{x}_k\right] \leq -1+\xi_k^+, \\ & \xi_k^-,\xi_k^+ \geq 0, \quad k=1,\ldots,m, \end{aligned} \tag{6}$$

ここで ξ_k^- と ξ_k^+ は制約条件を緩和するスラック変数で，パラメータ C によって調整される．この最適化問題をラグランジュ未定乗数法によって双対問題に変換したものを次式で示す．

$$\begin{aligned} & \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \sum_{k=1}^{m}(\alpha_k+\beta_k) - \frac{1}{2}\sum_{k,l=1}^{m} y_k^* y_l \lambda_k^* \lambda_l \langle \boldsymbol{x}_k\cdot\boldsymbol{x}_l\rangle \\ & \text{subject to :} \quad \lambda_k = \epsilon^*\alpha_k - \epsilon\beta_k, \\ & \quad 0\leq \alpha_k,\beta_k \leq C, \quad k=1\ldots,m. \end{aligned} \tag{7}$$

ここで α_k,β_k はラグランジュ未定乗数である．この二次計画問題を解いた後，CVSVM の出力を次式で求める．

$$s(\boldsymbol{x}) = \sum_{k=1}^{m} y_k\lambda_k \langle \boldsymbol{x}\cdot\boldsymbol{x}_k\rangle i. \tag{8}$$

3.2. カーネル法

カーネル法は式 (7) と (8) 中に出現するエルミート内積をカーネル関数で置き換える．これにより，高次元特徴空間での複雑なパターン分離が可能になる．本論文では以下のガウスカーネルを使用する．

$$K(\boldsymbol{x},\boldsymbol{z}) = \exp\left(-\left\|\boldsymbol{x}-\boldsymbol{z}\right\|^2/\sigma^2\right) \tag{9}$$

4. 計算機実験

計算機実験では，[2] から入手した 4 つのベンチマークを使って提案法の性能を比較した．各ベンチマークの性質を表 1 に示す．

比較手法として，提案法である CVSVM，従来法である MVN,MLMVN[1] を用いた．クラス分類は単一の分類器を用い，活性化関数 (2) を通した出力の偏角による分類を行った．また，CVSVM はカーネルを用いる場合と，用いない場合の両方を計測し，使用するカーネルとしては，式 (9) のガウスカーネルを用いた．

表 1: 計算機実験で使用したベンチマーク

Data Set	No. of Class	No. of Examples	No. of Attributes
Sonar	2	208	60
Ionosphere	2	351	34
Iris	3	150	4
Wine	3	178	13

なお，緩和項 $C=\{10^{-2},10^{-1},\ldots,10^4\}$ とカーネルパラメータ $\sigma=\{0.1,0.2,\ldots,4.0\}$ の最適な組み合わせは，10 分割の交差検定により決定した．MVN と MLMVN では，CVSVM と同じテストパターンと学習パターンを用いて，重みをランダムに初期化した１０回の学習結果の平均をとった．また，MLMVN では中間ニューロン数が 2 個，出力ニューロン数が 1 個の 3 層ネットワークを使用した．実験によって得られた平均認識率 (正しく分類できたテストパターンの割合の平均) を表 2 に示す．

結果より，カーネルを用いない場合の提案法 CVSVM(Linear) と従来法の MVN を比較すると，CVSVM の方が高い認識率を示している．また，カーネルを用いた場合 (RBF) でも，従来法の MLMVN よりも良い性能が得られている．このことから，提案法はマージン最大化による汎化性を得ており，未知のパターンに対する認識率が従来法よりも良くなる傾向があることがわかる．

表 2: データセットに対する平均認識率

Data	MVN	MLMVN	CVSVM (Linear)	CVSVM (RBF)
Sonar	0.798	0.792	0.865	0.914
Ionos.	0.864	0.881	0.920	0.957
Iris	0.951	0.945	0.987	0.987
Wine	0.915	0.938	0.960	0.980

5. まとめ

複素数の持つ性質を利用したニューラルネットワークである MVN を拡張し，新しい SVM の一種として CVSVM を提案した．提案法の CVSVM はマージン最大化法とカーネル法によって，従来のネットワーク型学習器よりも高い学習汎化性能を示した．今後は，より具体的な応用問題において，提案手法の性質がどのように活用できるかを検証していく．

参考文献

[1] I.Aizenberg and C.Moraga. Multilayer Feedforward Neural Network Based on Multi-Valued Neurons (MLMVN) and a Backpropagation Learning Algorithm. Soft Computing 11(2), pp.169-183.

[2] UCI machine learning repository. `http://archive.ics.uci.edu/ml/`