

非線形SVMによる属性の抽出

01703730 東京工業大学経営工学専攻 矢島 安敏 YAJIMA Yasutoshi
東京工業大学経営工学専攻 阿部 哲朗 ABE Tetsuo

1. はじめに

Support Vector Machine(SVM)を使った判別分析 [3, 1] の特徴の 1 つに, カーネル関数を用いた非線形な判別が行える点が挙げられる. しかし, SVM で得られた判別関数は判別の結果のみを与えるいわばブラックボックスであり, 例えば, 各属性の判別への寄与などを解釈することは困難である. そこで, SVM による判別に対し, Lee 等 [2] による判別境界を用いた属性抽出法を提案する. 本論文では, SVM による判別関数が特徴空間内では線形関数であることに注目し, データを判別境界面に直交射影することで境界上の点を生成する方法を提案する.

2. SVM による判別分析

N 個の属性を持った M 個のデータ $x_1, \dots, x_M \in \mathbb{R}^N$ およびクラスラベル $y_j \in \{-1, +1\}$ が与えられているとする. SVM では法線ベクトル $w \in \mathbb{R}^N$ と定数 $b \in \mathbb{R}$ で定まる線形判別関数 $f(x) = w^T x + b$ によりクラスの判別を考える. $f(x)$ を求めるために, $\alpha = (\alpha_1, \dots, \alpha_M)^T \in \mathbb{R}^M$ を変数とした以下の最適化問題

$$\begin{cases} \text{最小化} & -\sum_{j=1}^M \alpha_j + \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M \alpha_i \alpha_j y_i y_j x_i^T x_j \\ \text{制約} & \sum_{j=1}^M \alpha_j y_j = 0, \\ & 0 \leq \alpha_j \leq C, \quad j = 1, 2, \dots, M \end{cases}$$

を導入する. ここで, C は予め定められた正の実数である. α^* を最適解とすれば, $w = \sum_{j=1}^M y_j \alpha_j^* x_j$ となる関係を用いて w を求めることができる.

非線形な判別関数を構成するためには, \mathbb{R}^N の点をより高次の空間 \mathcal{F} へと写像する非線形変換 $\phi: \mathbb{R}^N \rightarrow \mathcal{F}$ を導入し, 空間 \mathcal{F} 上で点 $\phi(x_1), \dots, \phi(x_M) \in \mathcal{F}$ に対して線形関数 $G(\xi) = \langle w, \xi \rangle + b$ により判別を行う. 線形の場合を拡張して, 次の問題

$$(2.1) \quad \begin{cases} \text{最小化} & -\sum_{j=1}^M \alpha_j \\ & + \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M \alpha_i \alpha_j y_i y_j \mathcal{K}(x_i, x_j) \\ \text{制約} & \sum_{j=1}^M \alpha_j y_j = 0, \\ & 0 \leq \alpha_j \leq C, \quad j = 1, 2, \dots, M \end{cases}$$

を考える. ただし, $\mathcal{K}(x_i, x_j)$ は $\phi(x_i)$ と $\phi(x_j)$ との内積を与えるカーネル関数である. この問題の最適解を $\alpha^* \in \mathbb{R}^M$ とすれば, 最適な法線ベクトル $\omega^* \in \mathcal{F}$ は $\omega^* = \sum_{j=1}^M y_j \alpha_j^* \phi(x_j)$ と定められる.

3. 非線形 SVM における属性抽出法

3.1 非線形な判別境界と DBSM

以降では, 記号を単純にするため, 問題 (2.1) の最適解を単に $\alpha \in \mathbb{R}^M$, 等式制約に対応した最適な双対変数を $b \in \mathbb{R}$ と書き, 判別境界が入力空間で

$$B = \left\{ x \in \mathbb{R}^N \mid g(x) = \sum_{j=1}^M y_j \alpha_j K(x_j, x) + b = 0 \right\}$$

と求められたと仮定する. このとき, 以下の様に定義する [2].

定義 1 $v \in \mathbb{R}^N$ を大きさ 1 のベクトル, I を N 次単位行列として $U = I - vv^T$ を定める. 任意のデータ x_j ($j = 1, 2, \dots, M$) に対して, $g(x_j)g(Ux_j) \geq 0$ となる v を *discriminantly redundant* な方向と呼ぶ. 一方, あるデータ x_j に対し $g(x_j)g(Ux_j) < 0$ となるとき, v を *discriminantly informative* な方向と呼ぶ.

discriminantly redundant あるいは *informative* な方向を求めるために, 次の Decision Boundary Scatter Matrix (DBSM)

$$\Sigma_{DBSM} \equiv \int_B \frac{\nabla g(x) \nabla^T g(x)}{\|\nabla g(x)\|^2} p(x) dx$$

を考える. ここで, $p(x)$ はデータ x の確率密度関数とする. 今, 行列 Σ_{DBSM} の固有値をその大きさの順に $\lambda_1 \geq \dots \geq \lambda_N$ とし, それぞれに対応する大きさ 1 の固有ベクトルを $v_1, \dots, v_N \in \mathbb{R}^N$ とする. l を非 0 の固有値数とすれば, v_1, \dots, v_l で張る空間が判別に本質的な意味を持っている *discriminantly informative* な空間であり, また 0 の固有値に対応する固有ベクトル v_{l+1}, \dots, v_N で張られる部分空間が, *discriminantly redundant* な方向と考えることができる.

仮に, 十分な数のデータ $\hat{x}_1, \dots, \hat{x}_{M'}$ が判別境界上に得られていれば, 各点での法線ベクトル $\nabla g(\hat{x}_1), \dots, \nabla g(\hat{x}_{M'})$ を用いて, Σ_{DBSM} を

$$\hat{\Sigma}_{DBSM} \equiv \frac{1}{M'} \sum_{j=1}^{M'} \frac{\nabla g(\hat{x}_j) \nabla^T g(\hat{x}_j)}{\|\nabla g(\hat{x}_j)\|^2}$$

と近似的に求めることができる. しかし, 多くの場合境界上の点は極めて僅かで, 少ない境界上の点では十分によい近似が得られることは期待できない.

3.2 判別境界上の法線ベクトルの構成

本論文では、まず境界上の点を特徴空間 \mathcal{F} において生成し、これを用いて入力空間上で判別境界の法線ベクトルを求める方法を提案する。以降の節では、SVM によって作られた特徴空間上で判別関数を $g(\xi) = \mathcal{K}(\omega, \xi) + b$ と記すことにする。ただし、 $\alpha = (\alpha_1, \dots, \alpha_M)^T \in \mathbb{R}^M$ を問題 (2.1) の最適解として、 $\omega = \sum_{j=1}^M y_j \alpha_j \phi(x_j) \in \mathcal{F}$ と定まっているものとする。また、簡単のため各データ $x_k \in \mathbb{R}^N$ の非線形変換 ϕ による像を $\phi_k \equiv \phi(x_k) \in \mathcal{F}$ と書く。このとき、次の補題を得る。

補題 3.1 任意の点 $\phi_k \in \mathcal{F}$ ($k = 1, \dots, M$) に対して

$$\hat{\phi}_k \equiv \phi_k - \frac{\langle \omega, \phi_k \rangle + b}{\langle \omega, \omega \rangle} \omega$$

と定まる点 $\hat{\phi}_k \in \mathcal{F}$ は、 ϕ_k に最も近い境界上の点である。

次に、特徴空間での境界上の点 $\hat{\phi}_k \in \mathcal{F}$ に対して、

$$(3.1) \quad \phi(\hat{x}_k) = \hat{\phi}_k$$

となる入力空間の点 $\hat{x}_k \in \mathbb{R}^N$ を求める必要がある。しかし、SVM ではカーネル関数 $K(\cdot, \cdot)$ で内積のみを用いて判別境界を構成しているため、非線形変換 ϕ は陽に定められていない。したがって、一般的には (3.1) を満たす \hat{x}_k を求めることができず、さらには \hat{x}_k での法線ベクトル $\nabla g(\hat{x}_k)$ も求めることができない。しかし、以下に示すようにいくつかの代表的なカーネル関数の場合には、 \hat{x}_k を陽に求めることなく、 $\nabla g(\hat{x}_k)$ の計算が可能である。

まず、境界がカーネル関数の重ね合わせとして B として表現されているから

$$(3.2) \quad \nabla g(x) = \sum_{j=1}^M y_j \alpha_j \nabla K(x_j, x)$$

である。ここで、 $\nabla K(x_j, x)$ とは x_j を定数としたときの、変数 x に関する勾配ベクトルである。

補題 3.2 任意の 2 点 x_j, \hat{x}_k , ($j, k = 1, 2, \dots, M$) に対して

$$K(x_j, \hat{x}_k) = K(x_j, x_k) - \frac{\sum_{i=1}^M y_i \alpha_i K(x_i, x_k) + b}{\sum_{i, i'=1}^M y_i y_{i'} \alpha_i \alpha_{i'} K(x_i, x_{i'})} \sum_{i=1}^M y_i \alpha_i K(x_i, x_k)$$

である。すなわち、 $K(x_j, \hat{x}_k)$ の計算は \hat{x}_k を陽に求めることなく可能である。

補題 3.2 を用いれば、例えば、polynomial カーネルの場合には

$$\nabla K(x_j, x) = \kappa d \frac{K(x_j, x)}{\kappa x_j^T x + 1} x_j$$

である。特に、 d が奇数であれば、 $(\kappa x_j^T x + 1)^{d-1}$ は常に非負となることより、式 (3.2) を使い、

$$(3.3) \quad \nabla g(\hat{x}_k) = \kappa d \sum_{j=1}^M y_j \alpha_j |K(x_j, \hat{x}_k)|^{\frac{d-1}{d}} x_j$$

となる。補題 3.2 より、この式の右辺は \hat{x}_k を使わずに計算が可能である。

3.3 DBSM を用いた特徴の抽出

上の方法で求めた境界上の点 $\hat{x}_1, \dots, \hat{x}_M$ における法線ベクトルを使い行列 Σ を

$$\Sigma \equiv \frac{1}{M} \sum_{j=1}^M \frac{\nabla g(\hat{x}_j) \nabla^T g(\hat{x}_j)}{\|\nabla g(\hat{x}_j)\|^2}$$

とする。前節で述べたように、行列 Σ の非 0 の固有値 λ_i に対応した固有ベクトル v_i が、判別に本質的に寄与している方向である。そこで、本論文では固有値 λ_i の大きさを、方向 v_i の判別に寄与する度合と考えることにする。より一般的には、任意の方向 $v \in \mathbb{R}^N$ に対して、 v 方向の判別への寄与の度合い $F(v)$ を次のように定める。

定義 2 $v \in \mathbb{R}^N$ を大きさ 1 のベクトルとする。このとき、

$$(3.4) \quad F(v) \equiv v^T \Sigma v$$

を v 方向の判別への寄与度と定義する。

すなわち、任意の方向の中で、式 (3.4) を最大にするものがちょうど行列 Σ の第 1 固有ベクトルに対応する。また、 $e_i \in \mathbb{R}^N$ を第 i 方向の単位ベクトルとすれば、対応する第 i 属性の判別に寄与する度合いが $e_i^T \Sigma e_i$ となり、これは Σ の第 i 対角要素に他ならない。

なお、発表では、いくつかの数値実験を通じて、本論文の手法の有効性について報告する予定である。

参考文献

- [1] N. CRISTIANINI AND J. SHAWE-TAYLOR, eds., *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, U.K., 2000.
- [2] C. LEE AND D. A. LANDGREBE, *Feature extraction based on decision boundaries*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 15 (1993), pp. 388-400.
- [3] V. N. VAPNIK, *The nature of statistical learning theory*, Statistics for Engineering and Information Science, Springer-Verlag, New York, 2000.