日本オペレーションズ・リサーチ学会
2005年春季研究発表会

Conditional Geometric Scoreに基づく線形判別モデル

02103210 筑波大学 *後藤 順哉 GOTOH Jun-ya
01308490 東京工業大学 武田 朗子 TAKEDA Akiko

1. はじめに

90年代終盤以降, 金融リスク制御の分野において, Conditional Value at Risk (CVaR) が様々な観点から好ましいリスク尺度であるとして脚光を浴びている. 一方, データ・マイニングの分野において, サポート・ベクター・マシン (SVM) が実績を基に, 実務においても広くもてはやされている. 本研究では通常SVMの文脈において出発点として紹介されるhard-margin SVMを, CVaRのアイデアを適用したリスク尺度によって一般化し, その最小化に基づく線形判別モデルを提案する.

2. Conditional Geometric Scoreとその最小化

同一の分布から得られたラベル付データ $(\boldsymbol{x}^1, y^1), \ldots, (\boldsymbol{x}^\ell, y^\ell) \in \mathbb{R}^n \times \{\pm 1\}$ が所与であるとする. 線形判別モデルでは, これらのデータから適当な基準に基づいて判別超平面 $H(\boldsymbol{w}, b) := \{\boldsymbol{x} \mid \langle \boldsymbol{w}, \boldsymbol{x} \rangle + b = 0\}$ $(\boldsymbol{w} \neq \boldsymbol{0})$ を求める. データ $(\boldsymbol{x}^i, y^i)$ のgeometric score (margin) を以下で定義する:

$$g_i(\boldsymbol{w}, b) := \frac{y^i(\langle \boldsymbol{w}, \boldsymbol{x}^i \rangle + b)}{\|\boldsymbol{w}\|}.$$

また, 経験分布 $\mathrm{P}\{(\boldsymbol{x}, y) = (\boldsymbol{x}^i, y^i)\} = p_i > 0$ が既知であるとし, 固定された $(\boldsymbol{w}, b)$ に対し $-g(\boldsymbol{w}, b)$ の分布関数 Φ と β-quantile α_β を以下のように定義する:

$$\Phi(\alpha \,|\, \boldsymbol{w}, b) := \mathrm{P}\{-g(\boldsymbol{w}, b) \leq \alpha\},$$
$$\alpha_\beta := \alpha_\beta(\boldsymbol{w}, b) := \min\{\alpha \mid \Phi(\alpha \,|\, \boldsymbol{w}, b) \geq \beta\},$$

ただし, $\beta \in (0, 1)$. これらにより β-tail 分布関数を以下のように定義する:

$$\Phi_\beta(\alpha \,|\, \boldsymbol{w}, b) := \begin{cases} 0 & \text{for } \alpha < \alpha_\beta, \\ \frac{\Phi(\alpha | \boldsymbol{w}, b) - \beta}{1 - \beta} & \text{for } \alpha \geq \alpha_\beta. \end{cases}$$

β-tail 分布の下の期待値を $\phi_\beta(\boldsymbol{w}, b)$ と定義する. [2] によれば, 以下の関係が成り立つ:

$$\begin{aligned} \mathrm{E}[-g(\boldsymbol{w}, b) \mid -g(\boldsymbol{w}, b) \geq \alpha_\beta] &\leq \phi_\beta(\boldsymbol{w}, b) \\ &\leq \mathrm{E}[-g(\boldsymbol{w}, b) \mid -g(\boldsymbol{w}, b) > \alpha_\beta]. \end{aligned}$$

ここで, $\mathrm{E}[\cdot|\cdot]$ はもとの分布関数 Φ に基づく条件付き期待値である.

本研究では ϕ_β を最小化することによって判別超平面を求める. [2] によれば, この最小化は以下の最小化問題と等価である:

$$\left| \underset{\boldsymbol{w} \neq \boldsymbol{0}, b, \alpha}{\text{minimize}} \quad \alpha + \frac{1}{1 - \beta} \sum_{i \in I} p_i \, [-g_i(\boldsymbol{w}, b) - \alpha]^+ . \right. \tag{1}$$

ただし, $I := \{1, \ldots, l\}$, $[X]^+ := \max\{X, 0\}$. ここで次の非凸計画問題を考える:

$$(\mathrm{Q}(\beta)) \left| \begin{array}{ll} \underset{\boldsymbol{w}, b, \boldsymbol{z}, \alpha}{\text{minimize}} & \alpha + \frac{1}{1 - \beta} \sum_{i \in I} p_i \, z_i \\ \text{subject to} & (\boldsymbol{w}, b, \alpha, \boldsymbol{z}) \in H, \\ & \|\boldsymbol{w}\|^2 = 1. \end{array} \right. \tag{2}$$

ここで, $H := \{(\boldsymbol{w}, b, \alpha, \boldsymbol{z}) | z_i + y^i(\langle \boldsymbol{w}, \boldsymbol{x}^i \rangle + b) + \alpha \geq 0 \ (i \in I), \ \boldsymbol{z} \geq \boldsymbol{0}\}$ である.

Assumption 1 以下では次を仮定する:

- ▷ 各クラスが1つ以上のデータからなる.
- ▷ パラメータ β が以下を満たす:

$$1 - 2 \min\{ \sum_{i: y^i = 1} p_i, \sum_{i: y^i = -1} p_i \} \leq \beta < 1.$$

実際, Assumption 1の下で (1) と (2) は最適解を持つことが示される.

Proposition 2 $(\bar{\boldsymbol{w}}, \bar{b}, \bar{\alpha})$ を (2) の最適解とする. このとき, 任意の $\eta > 0$ に対し, $(\eta \bar{\boldsymbol{w}}, \eta \bar{b}, \bar{\alpha})$ は (1) の最適解となる.

3. 2-Step アルゴリズム

ここで, (2) の緩和問題を考える:

$$\left|\begin{array}{ll} \underset{w,b,z,\alpha}{\text{minimize}} & \alpha + \frac{1}{1-\beta}\sum_{i\in I} p_i z_i \\ \text{subject to} & (\boldsymbol{w}, b, \alpha, \boldsymbol{z}) \in H, \\ & \|\boldsymbol{w}\|^2 \leq 1. \end{array}\right. \quad (3)$$

Proposition 3 $(\hat{\boldsymbol{w}}, \hat{b}, \hat{\alpha}, \hat{\boldsymbol{z}})$ を (3) の最適解とする. このとき Assumption 1 の下で, 以下のいずれかが成り立つ (ただし, opt.$(Q(\beta))$ は (2) の最適値):

1. $\|\hat{\boldsymbol{w}}\| = 1$ ならば, $(\hat{\boldsymbol{w}}, \hat{b}, \hat{\alpha}, \hat{\boldsymbol{z}})$ は (2) の最適解になる. このとき, opt.$(Q(\beta)) \leq 0$.
2. $\|\hat{\boldsymbol{w}}\| \in (0,1)$ ならば, $\frac{1}{\|\hat{\boldsymbol{w}}\|}(\hat{\boldsymbol{w}}, \hat{b}, \hat{\alpha}, \hat{\boldsymbol{z}})$ は (2) の解であり, 最適値は 0 である.
3. $\|\hat{\boldsymbol{w}}\| = 0$ ならば, opt.$(Q(\beta)) \geq 0$ である.

Corollary 4 $(\hat{\boldsymbol{w}}, \hat{b}, \hat{\alpha}, \hat{\boldsymbol{z}})$ を (3) の最適解とする. このとき, 以下が成り立つ:

▷ opt.$(Q(\beta)) < 0$ ならば, $\|\hat{\boldsymbol{w}}\| = 1$.

▷ opt.$(Q(\beta)) > 0$ ならば, $\|\hat{\boldsymbol{w}}\| = 0$.

この結果を考慮し, 以下の 2 段階の枠組みを提案する:

Algorithm 1 (2-step framework)

Step 1. まず, 凸計画問題 (3) を解く. その最適解 $(\hat{\boldsymbol{w}}, \hat{b}, \hat{\alpha}, \hat{\boldsymbol{z}})$ に対し,

▷ $\|\hat{\boldsymbol{w}}\| = 1$ ならば, $(\hat{\boldsymbol{w}}, \hat{b}, \hat{\alpha})$ を (2) の解として終了.

▷ $\|\hat{\boldsymbol{w}}\| \in (0,1)$ ならば, $(\frac{\hat{\boldsymbol{w}}}{\|\hat{\boldsymbol{w}}\|}, \frac{\hat{b}}{\|\hat{\boldsymbol{w}}\|}, \frac{\hat{\alpha}}{\|\hat{\boldsymbol{w}}\|})$ を (2) の解として終了.

▷ $\|\hat{\boldsymbol{w}}\| = 0$ ならば Step 2 へ.

Step 2. 非凸計画問題 (2) の局所最適解を [4] のアルゴリズムにより求める.

4. ν-SVC との関係

ν-SVC は以下の問題を解く ([3]):

$$\left|\begin{array}{ll} \underset{w,b,z,\rho}{\text{minimize}} & \frac{1}{2}\|\boldsymbol{w}\|^2 - \nu\rho + \frac{1}{|I|}\sum_{i\in I} z_i \\ \text{subject to} & (\boldsymbol{w}, b, \boldsymbol{z}, \rho) \in N. \end{array}\right. \quad (4)$$

ただし, $N := \{(\boldsymbol{w}, b, \rho, \boldsymbol{z}) | z_i + y^i(\langle \boldsymbol{w}, \boldsymbol{x}^i\rangle + b) - \rho \geq 0 \ (i \in I), \ \boldsymbol{z} \geq \boldsymbol{0}, \ \rho \geq 0\}$. ここで,

$$\alpha = -\rho, \ \beta = 1 - \nu, \ p_i = \frac{1}{|I|} \ (i \in I) \quad (5)$$

とすると, (2) と (4) の関係が明らかになる.

Theorem 5 (2) の最適値が正のとき, パラメータを (5) によって置き換えた (4) は $\boldsymbol{w} = \boldsymbol{0}$ なる無意味な解を出力する.

Theorem 6 (2) の最適値が負のとき, パラメータを (5) によって置き換えた (2) と (4) は同じ判別超平面を解として持つ.

5. 計算実験

本研究で提示したモデルの振る舞いを示すため乳癌, 糖尿病, 肝臓癌の 3 つの 2 クラス判別用データに対して cross-validation を行った. その結果, 非凸性が本質となる状況において高い予測力を示す場合や, 他のモデルより優れた結果が得られた. 計算結果の詳細は発表時に示す. なお, 詳細については [1] を参照されたい.

参考文献

[1] J. Gotoh and A. Takeda, A linear classification model based on conditional geometric score, IPPS Discussion Paper Series No.1096 (2004) Univ.of Tsukuba.

[2] R.T. Rockafellar and S. Uryasev, Conditional value-at-risk for general loss distributions, *J. Bank. Finance* 26 (2002) 1443-1471.

[3] B. Schölkopf et al. New support vector algorithms, *Neural Comput.* 12 (2000) 1207-1245.

[4] A. Takeda and H. Nishino, On measuring the inefficiency with the inner-product norm in date envelopment analysis, *European J. Oper. Res.* 133 (2001) 377-393.