

最小二乗線形判別のクラス代表として 1-of-K 符号は万能ではない

関田 巖, 巽 久行

筑波技術大学 保健科学部 情報システム学科

要旨: 最小二乗線形判別法は, 線形判別分析法のように固有値問題を解く必要がなく, 逆行列問題だけで簡便に求められる利点がある。このとき目的ベクトルとなるクラス代表ベクトルに, 1-of-K 符号化が用いられることが多い。しかし, 一直線上に3クラス以上が位置するようなデータに対しては, 正確な識別を与えないことが知られている。本稿では, クラス代表ベクトルにクラス平均ベクトルを採用すれば, 最小二乗線形判別法でも線形判別分析と同様なよい識別結果が得られることを示す。

キーワード: 最小二乗線形判別, 1-of-K 符号, クラス平均, 目的ベクトル

1. はじめに

最小二乗線形判別法(以降 LSLC と呼ぶ)は, 量的データのクラス分類において, 線形判別分析法(以降 LDA と呼ぶ)に比べ, 固有値問題を解く必要がなく, 逆行列問題だけで簡便に求められる利点がある。ただし, 目的ベクトルに相当する近似先となるクラス代表ベクトルの選定に, 自由度がある。

クラス代表ベクトルとしては, しばしば, 1-of-K 符号化が用いられている[1-4]。このとき, LSLC で作られる写像が, ベイズ決定則での事後確率を線形近似するものになることが知られている[1][5]。

一方で, 各クラスのまとまりが一直線上に位置するような3クラス以上からなるデータに対しては, クラス代表ベクトルとして 1-of-K 符号化を用いて LSLC を行うと, 正しい識別が行えないことが知られている[2-4]。

本稿では, その改善策としてクラス内平均をクラス代表ベクトルとする方法を採用すれば, 一直線上に3クラス以上が位置するようなデータに対しても LSLC でよい識別結果が得られることを示す。さらに, 線形判別分析と同様な識別力を持つ見込みのあることも示す。実験結果により検証する。

2. 準備

2.1 最小二乗線形判別法 (LSLC)

入力データ(入力ベクトル)は, M 次元の列ベクトル $\mathbf{x} \in \mathbb{R}^M$ であり, K クラス $\{C_k\}_{k=1}^K$ のいずれかに属しているとする。クラスを代表する J 次元のクラス代表ベクトルを $\mathbf{c}_k \in \mathbb{R}^J$ とする。

\mathbf{x} から J 次元ベクトル $\mathbf{y} \in \mathbb{R}^J$ へのアフィン変換を線形変換で表現するために, \mathbf{x} を斉次座標で表現し,

$$\mathbf{x}^{\sim T} := [\mathbf{x}^T, 1],$$

$$\mathbf{y} = \mathbf{A}^T \mathbf{x}^{\sim},$$

とする。ここに, \mathbf{x}^T は \mathbf{x} の転置を表し, \mathbf{A} は $(M+1) \times J$ 次元の行列である。

各クラスの入力データをクラス代表ベクトルに集中させることを考え, クラス代表ベクトルとの最小二乗誤差 $Q := \sum_{k=1}^K \omega_k \mathbf{E}_{C_k} (\mathbf{y} - \mathbf{c}_k)^T (\mathbf{y} - \mathbf{c}_k)$ (1) が最小となるように \mathbf{A} を求める。

ここに, \mathbf{E}_{C_k} は, クラス C_k の範囲内で平均をとることを意味する。 ω_k は C_k の先見確率であり, (クラス C_k の入力データ数) / (全ての入力データ数) で近似する。

すると, \mathbf{A} は, 正規方程式

$$\mathbf{R} \mathbf{A}^T = \sum_{k=1}^K \omega_k \boldsymbol{\mu}_k^{\sim} \mathbf{c}_k^T$$

を解くことで求められる。ここに,

$$\boldsymbol{\mu}_k^{\sim} := \mathbf{E}_{C_k} \mathbf{x}^{\sim},$$

$$\mathbf{R} := \mathbf{E} \mathbf{x}^{\sim} \mathbf{x}^{\sim T}.$$

このとき, \mathbf{y} の張る空間は, $\{C_k\}_{k=1}^K$ を通る $K-1$ 次元超平面となっている[1]。

2.2 ベイズの決定則と 1-of-K 符号化

誤識別率最小となるベイズの決定則では, ベイズの事後確率 $p(C_k|\mathbf{x})$ ($k=1, \dots, K$) が最大となるクラスに \mathbf{x} を識別する。この事後確率からなる K 次元ベクトル

$$\mathbf{B}(\mathbf{x}) := [p(C_1|\mathbf{x}), p(C_2|\mathbf{x}), \dots, p(C_K|\mathbf{x})]^T$$

をアフィン変換を含めた線形変換 $\mathbf{D} \in \mathbb{R}^{(M+1) \times K}$ で最小二乗近似

$$\sum_{\mathbf{x}} (\mathbf{D}^T \mathbf{x}^{\sim} - \mathbf{B}(\mathbf{x}))^T (\mathbf{D}^T \mathbf{x}^{\sim} - \mathbf{B}(\mathbf{x})) \rightarrow \min$$

することを考える。すると \mathbf{D} は, 式(1)において, \mathbf{c}_k を 1-of-K 符号化 (k 次元目の要素だけ1でそれ以外の要素が0

のベクトル)

$$c_k = [0, \dots, 0, 1, 0, \dots, 0]^T$$

与えたときに、求められる A (ただし、 $J=K$) と一致する [1] [5]。

識別されるクラスは、 $D^T x$ の要素で最大値をもつ次元となる。これはまた、 $|c_k|$ ($k=1, \dots, K$) が一定値であるため、クラス代表ベクトルとの距離が最小のクラスに識別されることと同値である [1]。

2.3 1-of-K 符号化のクラス代表ベクトルでうまく最小2乗線形判別できない例

前述の 2.1, 2.2 より、クラス代表ベクトルとして、1-of-K 符号化がしばしば用いられている。しかし、各クラスが一直線上に並ぶようなときには、正しい線形決定境界が得られないことが知られている [2-4]。

一例を図1と表1に示す。図1における点が入力データであり、色でクラスが示されている。×が誤識別された入力データであり、×の色のクラスに誤識別されたことを表す。表1では、行が真のクラスで、列が識別されたクラスを表す。この例では、 C_2 がうまく識別されておらず、全体として 35/150 が誤識別されている。

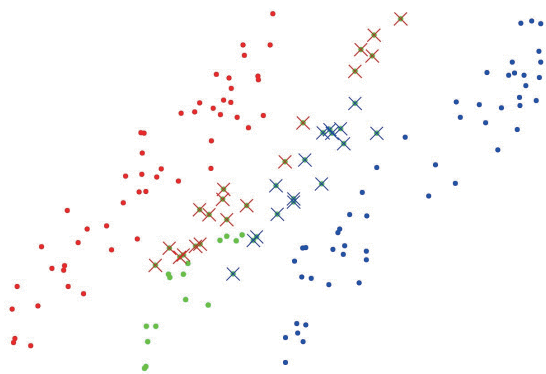


図1 直線上に並んだ3クラスに対し、1-of-K 符号化でLSLCをした結果 (×は誤識別のデータを示す)

表1 図1の識別結果表

	C_1	C_2	C_3
C_1	50	0	0
C_2	19	15	16
C_3	0	0	50

3. 1-of-K を使った最小2乗線形判別がうまくいかない場合の考察

複数 ($K \geq 3$) クラスが一直線上に並んでいるとき、入力データには線形従属性が高くなる。

一方で、1-of-K 符号化のクラス代表ベクトルは正規直交

系をなしているため、線形独立である。

線形従属なデータから線形独立なデータに変換するアフィン変換は存在しないが、1-of-K 符号化をクラス代表ベクトルとしてLSLCをすることは、線形従属なデータから線形独立なデータへのアフィン変換を最小2乗誤差基準で求めていることに相当する。このため、クラス間を分離する方向での広がり (直線状の広がり) に寄与しないクラス代表ベクトルが生じる。この寄与しないクラス代表ベクトルに対応するクラスの入力データは、正しく識別されなくなる。

4. 対策と有効性

クラス代表ベクトルとして、入力ベクトルのクラス内での平均 (クラス平均ベクトル)

$$c_k = E_{C_k} x = \mu_k$$

を採用し、そのときの有効性について考察する。

4.1 クラスが一直線上に並ぶデータに対して

クラスが一直線上に並ぶ入力データのモデルとして、各クラス C_k の入力データ ($x \in C_k$) が一定値

$$x = \mu_k (\forall k)$$

であり、 $\{\mu_k\}$ が一直線上に並んだ場合を想定する。そして、式 (1) を満たすアフィン変換 A^T について考察する。

クラス平均ベクトルをクラス代表ベクトルにしたときには、自明なアフィン変換

$$A^T = [I_M | 0]$$

が存在し、2乗誤差 $Q = 0$ となる。ここに、 I_M は、 $M \times M$ の単位行列で、 0 は M 次元の 0 ベクトルである。

一方、1-of-K 符号化をクラス代表ベクトルとするときには、3で示した通り、従属なベクトルを線形独立なベクトルに変換するアフィン変換は存在しないので、2乗誤差を0にすることができない。

4.2 線形判別分析に相当する判別基準

クラス平均ベクトルをクラス代表ベクトルにしたときには、各クラスの入力ベクトルがクラス平均ベクトルを近似するようにアフィン変換を求めているので、仮に、各クラス平均ベクトル μ_k が、アフィン変換により、近似的に

$$\mu_k$$

$$A^T \mu_k \approx \mu_k$$

と仮定する。すると、クラス間分散

$$\Sigma_B := 1/2 \sum_k \sum_l \omega_k \omega_l (\mu_k - \mu_l)(\mu_k - \mu_l)^T$$

は、 A^T によるアフィン変換後もほぼ不変

$$\Sigma_B \approx 1/2 \sum_k \sum_l \omega_k \omega_l (A^T \mu_k - A^T \mu_l)(A^T \mu_k - A^T \mu_l)^T$$

となる。

よって、クラス平均ベクトルをクラス代表ベクトルとする最小2乗基準 (1) は、

- (1) クラス間分散を一定にすること
 - (2) クラス内分散を最小化すること
- をできるだけ両立させるアフィン変換を求めているとみなせる。この基準は、(クラス内分散/クラス間分散)を最小化する判別分析の基準と同様な基準である。このため、判別空間上での距離に基づく識別と同様な識別力が得られることが期待される。

4.3 クラス代表ベクトルの役割分担

クラス平均ベクトルをクラス代表ベクトルにすることによって、クラス代表ベクトルの役割の中から、バイズの事後確率を近似するという役割を外すことができ、これにより柔軟性が高まり、線形従属な入力データに対してよりよい識別を与えるアフィン変換が得られるようになる。

クラス代表ベクトルは、各クラスとの対応ができている場合には、その対応関係はテーブルルックアップなどの非線形写像で容易に構築できる。さらに、未知入力ベクトルのアフィン変換後のベクトルと各クラス代表ベクトルからの距離に基づいて、所望する尺度（例えば事後確率に相当するものなど）を定義することもできる。

5. シミュレーション実験

5.1 クラスが一線上に並んでいる人工データ

前述の2.3と同じ入力データに対して、クラス平均ベクトルをクラス代表ベクトルとしてLSLCを適用した結果を図2と表2に示す。

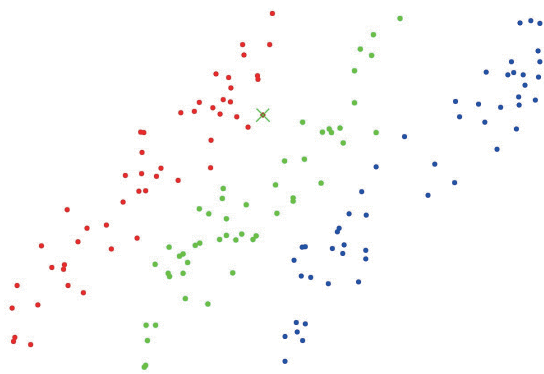


図2 直線上に並んだ3クラスに対し、クラス平均でLSLCをした結果(×は誤識別のデータを示す)

表1 図2の識別結果表

	C1	C2	C3
C1	49	1	0
C2	0	50	0
C3	0	0	50

図2における印の意味は図1と同様である。表2より、表1と比べて誤識別率が、35/150から1/150へと削減されており、クラス平均ベクトルをクラス代表ベクトルにする有効性を見ることができる。

5.2 フィッシャーのアイリスデータ

フィッシャーのアイリスデータに対して識別率を求めた。

1-of-K 符号化をクラス代表ベクトルにしたときの識別結果を表3に示す。クラス平均ベクトルをクラス代表ベクトルにしたときの識別結果を表4に示す。ここに、C1, C2, C3は、各々, setosa, visicolor, virginicaを表す。

表3 アイリスデータに対し、1-of-K 符号化でLSLCをした識別結果表

	C1	C2	C3
C1	50	0	0
C2	0	34	16
C3	0	7	43

表4 アイリスデータに対し、クラス平均でLSLCをした識別結果表

	C1	C2	C3
C1	50	0	0
C2	0	48	2
C3	0	1	49

クラス平均ベクトルをクラス代表ベクトルにすることにより、誤識別率が23/150から3/150に削減されることがわかる。

なお、表4の結果は、線形判別分析をした後に各クラスの分布が同一共分散行列を持つ正規分布と仮定して線形識別関数を構成したときの識別結果 [4]と同じであった。

6. おわりに

LSLCを適用する際、クラス代表ベクトルとして、1-of-K 符号化でうまく識別できない入力データに対しても、クラス平均ベクトルを採用することにより、識別できるようになることを示した。

この方法を用いれば、線形判別分析に必要な固有値問題を解かずに、逆行列問題を解くだけで、線形判別分析法と同等の識別結果が期待されることも示した。

今後の課題として、より厳密な理論的解析が期待される。

参照文献

- [1] 大津展之, パターン認識における特徴抽出に関する数理的研究, 電子技術総合研究所研究報告, 818, 1981.
- [2] C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
(cited 2020-09-25)
(<https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf>)
元田, 栗田, 樋口, 松本, 村田監訳: パターン認識と機械学習, 丸善出版
- [3] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, (Second Ed), Springer, 2009 (Revised 2017)
(cited 2020-09-25)
(https://web.stanford.edu/~hastie/ElemStatLearn/printings/ESLII_print12_toc.pdf)
- [4] 平井有三, はじめてのパターン認識, 森北出版, 2012.
- [5] W.G. Wee, Generalized inverse approach to adaptive multiclass pattern classification, IEEE Trans. Comp., C-17, pp.1157-1164, 1968.

1-of-K Coding for Target is Not Almighty in Least-squares Linear Classification

SEKITA Iwao, TATSUMI Hisayuki

Department of Computer Science, Faculty of Health Sciences,
Tsukuba University of Technology

Abstract: The least-squares linear classification method (LSLC) is more convenient than linear discriminant analysis (LDA) because the eigenvalue problem is not necessary. However, LSLC with a 1-of-K coding scheme for the target vector sometimes gives poor classifications especially when the number of class (K) is larger than 2 and the classes are mostly in line. This paper adopts the class mean vector for the target in LSLC, and shows that the LSLC gives almost the same classification performance as LDA.