共分散行列の違いを許容する線形判別分析

坂野 鋭[1,a)]　木村 昭悟[2]　藤木 淳[3]　田中 勝[3]

概要：Fisher の判別分析 (FLDA) はその簡便さ, 処理量の小ささ等から多くの研究者, 分析実務者をひきつけてきた．しかしながら FLDA は, すべてのクラスの分布が同じ共分散構造を持っているという強い仮定があるために，この仮定が満たされない場合には識別精度の低下を引き起こすという問題点があった．我々は，データの分布を包摂する部分空間同士を引き離すという考え方に基づいて，この問題を解決する詳細化判別分析を提案する．提案手法は FLDA の適用範囲を広げるのみならず FLDA の次元数の制約を取り払う．また, 処理量も大きくは無く，アルゴリズムも単純である．公開データによる実験的な評価の結果, 提案手法の有効性が明らかになった．

キーワード：次元圧縮, 特徴抽出, 判別分析, 部分空間法, CLAFIC

Novel Extended Fisher Discriminant Analysis

Abstract: Fisher's linear discriminant analysis (FLDA) has been attracting many researchers and practitioners for several decades thanks to its ease of use and low computational cost. However, FLDA implicitly assumes that all the classes share the same covariance, which implies that FLDA might fail when this assumption is not necessarily satisfied. To overcome this problem, we propose a simple extension of FLDA. The proposed method achieves remarkable improvements classification accuracy against FLDA while preserving two major strengths of FLDA: the ease of use and low computational costs. Experimental results with several data sets in UCI machine learning repository demonstrate the effectiveness of our method.

Keywords: Dimension Reduction, Feature extraction, Discriminant Analysis, Subspace Method, CLAFIC

1. はじめに

本論文において, 我々は Fisher の線形判別分析 (Fisher's Linear Discriminant Analysis; 以下 FLDA）を, クラス毎の共分散構造が異なる場合にも適用可能に拡張した詳細化判別分析 (Detailed Fisher's Discriminat Analysis; 以下 DFDA）を提案し, 実験的に有効性を示す.

Fisher の判別分析は [1]，恐らく最もよく知られた教師付次元圧縮法であり, その簡便さと性能の高さからパターン認識, コンピュータビジョン, 機械学習, 信号処理等の分野で幅広く用いられている．具体的な応用の中には文字認識や顔画像認識等の比較的身近なものもある.

1 (株)NTT データ
NTT Data, Koto, Tokyo 136–8671, Japan
2 NTT コミュニケーション科学基礎研究所
NTT Communication Science Lab.
3 福岡大学理学部応用数学科
Fukuoka University.
a) sakanoh@nttdata.co.jp

しかし一方で,FLDA はすべてのクラスの分散共分散行列が同一であるという仮定をしているために, この仮定が満たされない場合には著しい識別性能の低下を引き起こすことがある．また，級間分散の構造上, 削減された空間の次元数がクラス数－１次元に制限されるため, 少数クラス問題の場合には, やはり識別性能の低下を引き起こすことがある.

こうした問題を解決するためにこれまでに様々な研究が行われてきた．これらの研究は大きく二つに分けることができる．一つは（1）区分線形もしくは非線形の判別分析である．Hastie ら [8],Zhu ら [9],Gkalelis ら [10] は, それぞれ Gaussian Mixture を仮定した方法で区分線形近似を実現している．この方法はクラスタリングにより各クラスをほぼ同じ共分散構造を持ったサブクラスに分割することが期待される．一方,Baudat[6] や Sierra[7] は非線形変換により, 非ガウス性を克服することを目指した．これらの方法は有効な場合もあるが一般的には分布の混合数や非線形変

換のクラス等のモデル選択の問題を引き込むことがある．また，方法によっては局所最適解に落ち込む危険性もあり，改善の余地がある．

いま一つは (2) 級間分散の計算に辺り単純なユークリッド距離の代わりに Kullback-Leibler 情報量（以下,KL 情報量）や Chernoff 距離を用いたものである [13], [14]．このアプローチにより, クラス毎に共分散構造が異なる場合にも適用可能な判別分析法が構成されると期待できるが，現時点では KL 情報量や Chernoff 距離の非対称性や，ユークリッド空間の様な意味での平均が存在しないため, 定式化が複雑になりがちで適用範囲が事実上制限されるという問題がある．また，これらの研究の他に画像などの 2 次元の対象に適用するための改良もある [15].

本論文では，こうした現状を踏まえ，共分散構造がクラス毎に異なる場合にも適用可能な新しい判別分析法を提案する．提案法の基礎はクラス分布の広がりをクラスを包摂する部分空間で近似し, 部分空間同士の距離を広げることを試みることである [11]．以下, 次節では FLDA の問題点を具体化し，**3.** では提案手法の原理を示し，定式化を行う．**4.** において公開データセットを用いた評価実験結果を示す．**5.** において, 定式化及び実験結果についての考察を行う．**6.** はまとめと今後の課題である.

2. Fisher の線形判別分析とその問題

本節においては,FLDA の形式を示し, その問題点を明らかにする.

C クラスの識別問題を考える．$X_c = \{\vec{x}_1, \ldots, \vec{x}_{n_c}\}$ はクラス c に属する D 次元のサンプル集合である．ここで n_c はクラス c に属するサンプルの数である.

Fisher の判別分析は写像先で,C クラスの分布の分離がよくなることを目標とする．そのために級間分散行列

$$\Sigma_B = \frac{1}{C}\sum_{c=1}^{C}(\vec{\mu} - \vec{\mu}_c)(\vec{\mu} - \vec{\mu}_c)^\top, \tag{1}$$

及び，級内分散行列

$$\Sigma_W = \frac{1}{C}\sum_{c=1}^{C}\Sigma_c, \tag{2}$$

に随伴する 2 次形式をそれぞれ最大化, 最小化する写像を求める．ここで $\vec{\mu}$ は全サンプルの平均ベクトル，$\vec{\mu}_c$ はクラス c に属するサンプルの平均ベクトル, Σ_c はクラス c に属するサンプルの分散共分散行列

$$\Sigma_c = \frac{1}{n_c}\sum_{i=1}^{n_c}(\vec{\mu}_c - \vec{x}_{ci})(\vec{\mu}_c - \vec{x}_{ci})^\top. \tag{3}$$

である.

この問題は, 一般化固有値問題

$$\Sigma_B\vec{a} = \Sigma_W\vec{a}\lambda, \tag{4}$$

として解くことができる．ここで $\vec{a}$, λ は，それぞれこの一般化固有値問題の, 固有ベクトル，固有値である．この固有ベクトル $\vec{a}$ が写像先の空間の基底となる.

式 (1) からわかるとおり，行列 Σ_B のランクは $C-1$ であり，圧縮される次元数はこの制限を受ける．即ち,FLDA によって得られる写像後の空間の次元数は $D < C-1$ の場合は D 次元, そうではない場合には $C-1$ 次元が最大となる.

例えば $D = 1000$ の場合でも 2 クラス問題であれば 1 次元の特徴しか得られない．このとき，クラス毎の共分散構造が異なる場合には深刻な識別精度の低下を引き起こす可能性がある.

また,FLDA では, クラス毎の分布を表現する級内分散について, 式 (2) に見られるようにクラス毎の分散共分散行列を平均してしまい，また，クラスの違いを表す級間分散行列は式 (1) で判るとおり，各クラスの平均ベクトルのみから計算される．つまり, 各クラスの分布の違いに関する情報は, ほぼ完全に無視される．無論, 分布の違いは識別のために重要な情報であり, これを無視したことにより，識別精度が低下することは十分に考えられるし，現実にもそのような現象はよく見られる.

こうした問題に関する最も直接的なアプローチは分布間距離として, 単純なユークリッド距離の代わりに KL 情報量や Chernoff 距離の様に, クラス毎の共分散構造を反映した距離尺度を使うことが考えられる [13], [14]．こうした改良により, クラス毎の分布の情報を加味された写像が得られる．しかし,KL 情報量にせよ Chernoff 距離にせよユークリッド空間における平均ベクトルに対応する量が無いため，級間分散はクラス間のペアワイズ距離行列として定義されなくてはならない. 従って級間分散行列と級内分散行列の次元が異なり, かつ漢字認識問題の様な多クラス問題ではスケーラビリティの問題が生じる．また，これらの分布間距離は非対称であるために定式化の上でも疑問が残る.

3. 詳細化判別分析

本節において, 我々はクラス毎の共分散構造の違いを反映した判別分析，DFDA を提案する.

我々が提案する新しい判別分析 DFDA は CLass Featureing Information Compression(以下,CLAFIC) に基礎を置いている [12], [16]．CLAFIC は分布の表現として, サンプルが属する部分空間を用いる．識別時には各クラスを表現する部分空間に識別対象サンプルが所属するかどうかを判定する.

これらのクラス c を表現する部分空間の k 番目の基底 $\vec{\psi}_{ck}$ はクラス c の学習サンプルから計算された自己相関行列

$$\Gamma_c = \frac{1}{n_c}\sum_{\vec{x}\in\omega_c}\vec{x}\vec{x}^\top. \tag{5}$$

の固有ベクトルとして与えられる．$\vec{\psi}_{ck}$ は分散共分散行列の固有ベクトルではないが，明らかにそのクラスの分布情報を反映している．実際, クラス自己相関行列とクラス分散共分散行列の間には

$$\Sigma_c = \Gamma_c - \vec{\mu}\vec{\mu}^\top \tag{6}$$

という関係があり, クラス平均の情報を除けば, 分布の広がりに関する情報は自己相関行列の中にも保存されていると考えてよい．容易にわかる通り, サンプルの平均が原点と一致するときは自己相関行列と分散共分散行列は一致する．

我々の基本的なアイデアは, このクラスを包摂する部分空間同士を引き離すような写像を構成することである．直感的には $\vec{\psi}_{ck}$ を他のクラスの $\vec{\psi}$ から引き離すことでこれは実現される．従って，我々の提案する新しい評価関数は

$$\Sigma_{B2} = \sum_c \sum_{d \neq c} \sum_{k=1}^{d_u} \sum_{l=1}^{d_u} (\psi_{ck} - \psi_{dl})(\psi_{ck} - \psi_{dl})^\top. \tag{7}$$

のように表現される．ただし，$\vec{\psi}_{ck}$ は正規化されているものとする．ここで d_u は使用される自己相関行列の固有ベクトルの数である．

この評価関数の元で新しい級間分散行列は

$$\Sigma_{B_{new}} = \Sigma_B + \Sigma_{B2}. \tag{8}$$

のように定義される．明らかにこの行列は $C-1$ より大きなランクを持ち，かつクラス毎の分布の情報を保持しているため, 識別率の向上が期待される．

4. 実験的評価

本節において，我々は公開データセットによる評価実験結果を通して提案手法である DFDA の有効性を示す．評価のためには UCI Machine Learning repository[*1] から, 識別問題のために用意されたデータベースのうち，クラス数が次元数を下回るものを選択した．

判別分析はそれ自身が識別器として用いられることも多いが，今回の評価では特徴抽出器としての性質を把握することを重視し，再現性の観点から 1-最近傍識別器を用いた．評価結果を表 1. に示す．

先述したとおり，FLDA による次元圧縮の結果, 元空間のそれより識別精度が低下している例は多く見られる．DFDA の識別精度は今回の実験の範囲ではすべて FLDA を上回っており, 12 件中 9 件では元空間の識別精度を超えている．この結果から提案手法の有効性は明らかであろう．

表 1. の結果から DFDA で識別精度が高くなる圧縮空間の次元数が FLDA よりかなり大きいことがわかる．この結果は FLDA では少数の特徴しか抽出されないが, その識別能力は高く，DFDA では多数の特徴が抽出される反面, その識別能力は低いことが推測される．そこで,DFDA と FLDA で抽出された特徴を統合することを試みた．表 1 の Fusion と表示した項目にその結果を示す．統合は,FLDA,DFDA で圧縮された空間でのユークリッド距離の輪をとることで実現している．この結果から元空間での識別精度に対し DFDA では改善できない場合でも改善される場合があることがわかる．

*1 http://archive.ics.uci.edu/ml

5. 考察

処理量について

DFDA は, クラス毎に自己相関行列の計算及び対角化を行う必要があるため, FLDA に比べておおよそ $C+1$ 倍の処理量が必要となる．数クラスの問題に対しては, この程度の計算時間の増大は現代の計算機環境では大きな問題にはならないと考える．一方, 漢字認識問題の様にクラス数が数千に及ぶ多クラスの場合には問題となりうるが, 自己相関行列の対角化操作はクラス毎に独立で行えるため, 容易に並列化できる．従って，スケールアウトの際にも致命的な問題にはなりにくいと考えられる．

適用可能性について

評価に用いたデータセット 12 件のすべての場合について DFDA の識別精度は FLDA のそれを超えており, DFDA の適用範囲が FLDA のそれより広いことは明らかである．しかし，一方で Breast cancer, magic, statlog(Landsat) データでは元空間の識別精度を下回る結果となっており DFDA の適用範囲には限界があることを示していると考えられる．また，このような場合でも FLDA との統合により，さらに識別精度が向上する場合があることは, 提案手法には正統的な改善の余地があることを示唆していると考えられる．

原点について

FLDA は平行移動に関して不変な構造をしているが, DFDA は原点を共有する部分空間による記述を基礎としているため, 平行移動に関して不変とはいえない．直感的には全データの重心を原点に用いるのが良好そうであり, 今回の評価実験でもその様に原点を選んでいる．しかし, 将来的にデータの分離がよくなるように原点を選択するといった研究もありうるかもしれない．

確率密度関数との関係

DFDA においては, 次元圧縮と確率密度関数の関係が必ずしも明らかではない．これは Gauss 分布の密度関数が分散共分散行列の逆行列で表現されるために, 自己相関行列との関係を明示できないことに起因する．この問題に対して Fujiki は DFDA がクラス自己相関行列の固有値の判別分析と等価であることを示している．このことが確率密度関数の問題と関係するかもしれない [18].

6. まとめと今後の課題

本論文においては，クラス毎に共分散構造が異なる場合にも適用可能な新しい判別分析法 DFDA を提案し, 実験的

に有効性を示した．提案手法は簡単に実装出来, 処理量も大きくは無く, 多くの場合に識別精度の向上が見られる実用的な方法である．

しかし, 一方で明らかに適用範囲には限界がありこれを明らかにすることで更なる改良を行う可能性が高いと考えられる．

参考文献

[1] R. A. Fisher, The use of multiple measurements in taxonomic problems. Annals of Eigenics, 7:179-188, 1936

[2] R. O. Duda and P. E. Hart, Pattern Classification and Scene Analysis, John Willey and Sons, 1973

[3] P. N. Belhumeur, et al, "Eigenfaces vs Fusherfaces: Recognition using class specific linear projection," IEEE Transaction of Pattern analysis and Machine PAMI, Vol. 19, pp. 711-720, 1997

[4] T. Hastie, A. Buja and R. Tibshirani, "Penelized Discriminant Analysis," The Annals of Statistics, Vol. 23, No. 1 pp. 73-102, 1995

[5] K. Fukunaga, "Introduction to Statistical Pattern Recognition 2nd ed," Academic Press, 1990.

[6] G. Baudat and F. Anouar, "Generalized Discriminant Analysis Using a Kernel Approach," Neural Computation, Vol. 12, N0. 10, pp.2385-2404, 2006

[7] A. Sierra, "High-order Fisher's discriminant analysis," Pattern Recognition, Vol. 35, no. 6, pp. 1291-1302, 2002.

[8] T. Hastie and R. Tibshirani, "Discriminant Analysis by Gaussian Mixture," J. Royal Society of Statistical. Soc. B., Vol 58, pp. 155-176, 1996

[9] M. Zhu and A. M. Martinez, "Subclass discriminant analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 8, pp. 1274-86, Aug. 2006.

[10] N. Gkalelis, V. Mezaris, and I. Kompatsiaris, "Mixture subclass discriminant analysis," IEEE Signal Processing letters, Vol. 18, no. 5, pp. 319-322, 2011.

[11] H. Sakano, T. Ohashi, A. Kimura, H. Sawada and K. Ishiguro, "Extended Fisher Criterion Based on Autocorrelation Matrix Information," In Proceedings of Structural, Syntactic, and Statistical Pattern Recognition - Joint IAPR International Workshop, SSPR and SPR 2012, pp. 409-416, 2012.

[12] H. Sakano, "A Brief History of Subspace Method," ACCV2010 Workshop on Subspace 2010, 2010

[13] H.P. Decell and S.M. Mayekar, "Feature Combinations and the Divergence Criterion," Computers and Math. with Applications, vol. 3, pp. 71-76, 1977.

[14] M. Loog and R. P. W. Duin, "Linear dimensionality reduction via a heteroscedastic extension of LDA: the Chernoff criterion.," IEEE transactions on pattern analysis and machine intelligence, vol. 26, no. 6, pp. 732-9, Jun. 2004.

[15] T. Yoshida and Y. Yamada, "Toward Robust and Fast Two-Dimensional Linear Discriminant Analysis, " In Proceedings of Active Media Technology, pp.126-135, 2013.

[16] S. Watanabe, P.F. Lambert, C.A. Kulikowski, J.L. Buxton, and R. Walker, Evaluation and selection of variables in pattern recognition, Comp. & Info. Sciences, Vol. 2, (Julius Tou, ed.). New York: Academic Press, pp. 91–122, 1967.

[17] e.g. G. Lim and C. H. Park, "Semi-supervised Dimension Reduction Using Graph-Based Discriminant Analysis," in 2009 Ninth IEEE International Conference on Computer and Information Technology, 2009, pp. 9-13.

[18] J. Fujiki, "Finding discriminant axes from multiple viewpoints," In Proc. the 13th IAPR International Conference on Machine Vision Applications, pp. 73-76, 2013.

表 1 Evaluation of proposed method on UCI Machine Learning Repository data. "OFS" means the classification accuracy for the original feature space. "Fusion" is the classification accuracy when features extracted by FLDA and DFDA are fused. The total number of training samples are shown at the column of $n_c \times C$.r means dimensionality of extracted feature.

Data	D	C	$n_c \times C$	OFS	FLDA	DFDA	r, d_u	Fusion	r, d_u
Breast cancer	28	2	200	94.8%	78.0%	89.4%	26,13	**94.9%**	14,28
magic	10	2	200	71.3%	55.8%	70.9%	8,4	**71.8%**	3,6
wine	13	3	60	69.5%	40.0%	**91.5%**	13,6	87.3%	6,13
spambase	57	2	200	63.8%	55.8%	**83.8%**	56,2	72.3%	57,57
image segmentation	19	7	210	87.7%	48.9%	88.4%	19,18	**88.6%**	10,19
ionosphere	34	2	100	84.9%	56.6%	89.6%	34,24	**90.4%**	18,34
statlog(Landsat)	36	6	1800	**77.0%**	26.1%	76.4%	36,35	74.8%	6,36
statlog(Shuttle)	9	7	43500	**99.8%**	91.4%	**99.8%**	9,9	99.6%	9,9
statlog(viecle)	18	4	400	59.9%	37.2%	**70.4%**	18,15	67.7%	6,18
madelon	500	2	2000	75.7%	54.2%	**77.7%**	500,477	70.8%	367,500
optdigits	64	10	3823	97.8%	45.4%	**98.2%**	64,56	92.4%	18,64
Cardiotocographic	21	3	1000	81.3%	73.5%	**87.9%**	19,6	79.3%	7,21