AIMS Mathematics, 11(8): 26168–26185.
DOI:10.3934/math.20261048
Received: 06 June 2026
Revised: 30 July 2026
Accepted: 12 August 2026
Published: 24 August 2026

Mathematics

https://www.aimspress.com/journal/Math

Research article

Geometric partitioning-based multimodal linear discriminant analysis

Xinyi Han and Chaojie Wang*

School of Mathematics and Statistics, Beijing Technology and Business University, Beijing 100048, China

* **Correspondence:** Email: wangcj2019@btbu.edu.cn.

Abstract: Multimodal linear discriminant analysis(LDA), as an important method for cross-modal dimensionality reduction and feature fusion, has been extensively studied and applied in the field of pattern recognition. In this paper, a geometric partitioning-based multimodal linear discriminant analysis (GMLDA) method is proposed. The hyperplanes are constructed to partition the multimodal data within each class. The kernel principal component directions are used to determine the normal vectors of the hyperplanes, and they are served as the intra-class partitioning directions. The key point of the proposed method is the combination of hyperplanes, which makes the sample distribution within each modality more compact while increasing the separation between different modalities. Numerical results demonstrate the efficiency of the proposed GMLDA method by comparison with other multimodal LDA methods.

Keywords: linear discriminant analysis; multimodality; hyperplane; geometric partitioning; dimensionality reduction
Mathematics Subject Classification: 15A18, 62H30, 68T10

1. Introduction

Dimensionality reduction is a key technique for handling high-dimensional data, which can effectively alleviate the curse of dimensionality and improve computational efficiency. Linear discriminant analysis (LDA) is a highly representative method among them. Its core principle is to seek the optimal projection directions that minimize the intra-class scatter of samples in the same category and maximize the inter-class scatter of samples in different categories after projection, thereby reducing data dimensionality while preserving class-discriminative information [1]. LDA boasts prominent interpretability, strong robustness to noise, and a wide range of applicability, and thus has been extensively applied in numerous fields such as face recognition and image classification.

However, the LDA method has certain limitations. In recent years, a series of improvements

have been made to LDA. For instance, the robust sparse linear discriminant analysis (RSLDA) method [2] enhances the robustness and sparsity of LDA. To address the singularity problem of LDA, Wang et al. proposed the trace ratio LDA (TLDA) method [3]. The class and cluster-based LDA (ccLDA) method [4] alleviates the small sample size problem of LDA. Nevertheless, a large number of non-Gaussian and multimodal datasets exist in real-world scenarios [5]. Their categories are usually composed of multiple mutually independent subclasses or clusters, that is, they exhibit multimodal characteristics, rendering traditional LDA-based methods no longer applicable. To tackle the challenges posed by multimodal data, researchers have conducted extensive studies. Local Fisher discriminant analysis (LFDA) [6] focuses on discovering the local structure of data and performs LDA on the local region surrounding each data point. Neighborhood linear discriminant analysis (NLDA) [7] constructs neighborhoods using the reverse k-nearest neighbor (KNN) method and optimizes projection directions with the objective of minimizing intra-neighborhood scatter and maximizing inter-neighborhood scatter. Manifold partition discriminant analysis (MPDA) [8] introduces a linear embedding space where intra-class similarity is achieved along directions consistent with the local variations of the data manifold, and adjacent data points belonging to different classes are well separated. Mixture discriminant analysis (MDA) [9] serves as a classic solution to handle multimodal classes. Zhu and Martinez [10] introduced Gaussian mixtures to model subclasses for MDA, yet the count of mixture components cannot be easily decided. Gkalelis et al. [11] thus presented an iterative strategy to estimate component numbers, terminating iterations once the non-Gaussian measurement converges to a tiny value. For retaining local information in MDA, Tao et al. [12] combined k-means and an expectation-maximization (EM) -like framework to build a two-step subclass partition approach, which also fails to avoid the challenge of unknown cluster numbers. In addition, Zhu and Martinez [10] redefined the inter-class scatter matrix based on subclasses, and Chumachenko et al. [13] adopted graph embedding and spectral regression to speed up subclass discriminant analysis. Zhang et al. [14] proposed the TRANS-CNN model, which adopts transformer to capture long-range dependencies embedded in point cloud sequences and further leverages convolutional neural networks (CNNs) to extract local motion features. Besides, other methods based on auto weight [15], separability-oriented subclass [16], and variable-weighted separability-oriented subclass [17] have also been investigated to improve the efficiency of the LDA method for dealing with multimodal data. Although the aforementioned methods have improved the adaptability to multimodal data to a certain extent, most of them still rely on pairwise sample similarity evaluation or KNN strategies, leading to high computational complexity, and there remains room for improvement in the accuracy of multimodal structure recognition.

In pattern recognition and machine learning, hyperplanes are fundamental yet powerful geometric concepts, often used as decision boundaries for classification. Geometrically, a hyperplane is an $N-1$-dimensional subspace within an N-dimensional space, dividing the space into two disjoint half-spaces. This makes hyperplanes a natural ideal tool for linear discrimination and classification. Introducing the hyperplane mechanism into LDA is expected to enable it to actively identify intra-class multimodal structures and learn the most discriminative local projection directions for each sub-modality, thereby fundamentally enhancing the model's ability to characterize multimodal data. This paper proposes a new supervised dimensionality reduction algorithm—geometric partitioning-based multimodal linear discriminant analysis (GMLDA). This algorithm employs a geometric approach (hyperplane partitioning) to segment intra-class multimodal data; the segmented sub-classes

then replace the original classes in LDA for dimensionality reduction. Moreover, the direction of segmentation is adaptively determined via kernel principal component analysis (KPCA). Compared with existing methods, GMLDA has several notable advantageous properties: It segments intra-class multimodality through geometric partitioning, and uses KPCA to identify the nonlinear direction of multimodal segmentation during the process, which allows for more effective division of intra-class multimodal data with complex distributions. Compared with existing methods, GMLDA has several noteworthy advantages:

1) GMLDA partitions intra-class multimodality via geometric partitioning.
2) During segmentation, GMLDA uses KPCA to identify the directions for partitioning multi-modal structures, allowing for effective separation of intra-class multi-modalities with nonlinear characteristics.
3) GMLDA no longer relies on KNN or Euclidean distance for segmentation, which can significantly reduce computational complexity.

The remainder of this paper is organized as follows. In Section 2, a brief description of the LDA method and the hyperplane partitioning is given. In Section 3, the GMLDA method is presented. Numerical results are shown in Section 4. Section 5 draws the conclusions.

2. Background

This section briefly introduces the LDA model and the basic theory of hyperplane partitioning. We first clarify the notations used in this paper: Let the training set be $X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{m\times n}$, where each sample $x_i \in \mathbb{R}^m$ is a column vector with the corresponding class label $y_i \in \{1, 2, \ldots, c\}$ (c denotes the number of classes). Denote the l-th class sample set as X_l, with sample size n_l, class mean μ_l, and overall mean μ; the c-th class sample set is expressed as: $\mathcal{X}_c = \{x_i \mid y_i = c\}$.

2.1. LDA

LDA is a classic supervised dimensionality reduction method, whose goal is to find a linear projection matrix $\mathbf{W}$ such that the projected space achieves maximizing inter-class separability and intra-class compactness.

The within-class scatter matrix $\mathbf{S}_w$ defined by LDA reflects the degree of dispersion of intra-class samples around their class centroids, and it is desirable to minimize this scatter after projection. The between-class scatter matrix $\mathbf{S}_b$ characterizes the separation degree among the centroids of different classes, with the expectation to maximize this scatter following projection. Their calculation formulas are given respectively as follows:

$$\mathbf{S}_w = \sum_{l=1}^{c} \sum_{\mathbf{x}_i \in X_l} (\mathbf{x}_i - \boldsymbol{\mu}_l)(\mathbf{x}_i - \boldsymbol{\mu}_l)^T, \tag{2.1}$$

$$\mathbf{S}_b = \sum_{l=1}^{c} n_l (\boldsymbol{\mu}_l - \boldsymbol{\mu})(\boldsymbol{\mu}_l - \boldsymbol{\mu})^T. \tag{2.2}$$

Its objective function can be written in the form of a generalized rayleigh quotient:

$$J(\mathbf{W}) = \frac{|\mathbf{W}^T\mathbf{S}_b\mathbf{W}|}{|\mathbf{W}^T\mathbf{S}_w\mathbf{W}|}. \tag{2.3}$$

The optimal projection matrix of LDA is given by the following generalized eigenvalue decomposition problem:

$$\mathbf{S}_b\mathbf{W} = \lambda\mathbf{S}_w\mathbf{W}. \tag{2.4}$$

By solving the above equation, we can obtain the matrix $\mathbf{W}^*$ composed of the eigenvectors corresponding to the first m largest eigenvalues, which is the optimal projection direction.

2.2. Hyperplane partitioning

A hyperplane in $\mathbb{R}^m$ is an affine subspace of co-dimension one, which can be expressed as:

$$\mathbf{w}^\top\mathbf{x} + b = 0, \tag{2.5}$$

where $\mathbf{w} \in \mathbb{R}^m$ is the normal vector that determines the orientation of the hyperplane, and $b \in \mathbb{R}$ is the bias term that controls its position. The hyperplane splits the ambient space into two disjoint half-spaces: a point $\mathbf{x}$ satisfies $\mathbf{w}^\top\mathbf{x} + b > 0$ on one side and $\mathbf{w}^\top\mathbf{x} + b < 0$ on the other. By constructing multiple hyperplanes, the data space can be hierarchically partitioned into distinct regions. The application of this geometric tool to intra-class partitioning is developed in Section 3.1.

3. The proposed method

The central idea of GMLDA is to replace the clustering-based subclass assignment prevalent in existing multimodal LDA methods with a geometric operation: partitioning the feature space by hyperplanes. A hyperplane is an explicit geometric object—an affine subspace of co-dimension one that divides the ambient space into two disjoint half-spaces. By constructing hyperplanes within each class, GMLDA carves the intra-class region into compact subclasses separated by hard geometric boundaries, rather than assigning samples to subclass centers by proximity.

Sections 3.1–3.4 develop this idea systematically: Section 3.1 formalizes the hyperplane partitioning operation itself—this is the core methodological contribution. Section 3.2 addresses a necessary operational question within this framework—how to orient each hyperplane—and provides theoretical justification for using KPCA as the solution. Section 3.3 presents a data-driven rule for determining how many hyperplanes to construct. Section 3.4 assembles the complete GMLDA algorithm by combining hyperplane-partitioned subclasses with the LDA objective.

3.1. Intra-class partitioning via hyperplanes

Consider class c with sample set $\mathcal{X}_c = \{\mathbf{x}_i \mid y_i = c\} \subset \mathbb{R}^m$, where $|\mathcal{X}_c| = N_c$. In multimodal scenarios, $\mathcal{X}_c$ comprises K_c distinct sub-modalities—compact clusters separated by low-density regions in the feature space. The goal is to decompose $\mathcal{X}_c$ into K_c subsets $\{\mathcal{X}_{c,k}\}_{k=1}^{K_c}$, each corresponding to a single modality.

Existing subclass-based LDA methods approach this as a clustering problem: Each sample is assigned to the subclass whose centroid is closest in Euclidean distance, and the centroids are iteratively

refined to optimize a separability or compactness criterion [4,16]. The resulting subclass boundaries are implicit—they are the Voronoi cells induced by the subclass centroids, without any explicit geometric structure delineating one subclass from another.

GMLDA takes a fundamentally different route. Instead of clustering samples around centroids, we partition the space directly with hyperplanes. As defined in Section 2, a hyperplane (Eq (2.5)) is parameterized by a normal vector $\mathbf{w}$ and a bias b, and deterministically splits the space into two half-spaces. This geometric formulation leads to the following formal definition of intra-class partitioning.

Definition 3.1 (Hyperplane partition of a class)**.** *For class c, we construct* K_c-1 *hyperplanes* $\{\mathcal{H}_{c,k}\}_{k=1}^{K_c-1}$, *where*

$$\mathcal{H}_{c,k} : \mathbf{w}_{c,k}^{\top}\mathbf{x} + b_{c,k} = 0. \tag{3.1}$$

Each hyperplane $\mathcal{H}_{c,k}$ applies a binary split to the subclass it operates on, dividing it into two child subclasses according to the sign of $\mathbf{w}_{c,k}^{\top}\mathbf{x} + b_{c,k}$. After $K_c - 1$ successive splits, class c is partitioned into K_c subclasses.

Hyperplane formulation confers three structural advantages over clustering-based partitioning. First, geometric interpretability: Each hyperplane is an explicit boundary—one can inspect its normal vector to understand which direction separates the sub-modalities and its bias to see where the separation occurs. Second, deterministic assignment: A sample's subclass membership is determined purely by which side of each hyperplane it falls on, without iterative optimization or distance computations. Third, natural parameterization: A hyperplane is parameterized by a direction vector $\mathbf{w}$ and a scalar b, making it amenable to principled estimation (Sections 3.2 and 3.3) rather than heuristic specification.

Two key practical issues require resolution for the hyperplane framework. The first issue concerns the orientation of each hyperplane. The normal vector $\mathbf{w}_{c,k}$ must align with the direction of separation between sub-modalities, otherwise the cut cannot isolate distinct clusters. The second issue concerns the total number of hyperplanes. An insufficient number of hyperplanes fails to separate distinct sub-modalities, while excessive hyperplanes lead to data over-partitioning and introduce singularity risk. Sections 3.2 and 3.3 address these two issues in sequence.

3.2. Orienting the hyperplanes: normal vector determination

For a hyperplane to separate two sub-modalities, its normal vector $\mathbf{w}$ must point along the direction in which those sub-modalities are maximally apart. Intuitively, within a multimodal class, the principal directions of variation—the directions along which the data spread the most—should reveal the axes of inter-modality separation, since the displacement between distinct clusters contributes dominantly to the overall variance.

KPCA provides a principled means of identifying these directions while capturing nonlinear structure. In what follows, we formalize this intuition and provide theoretical guarantees.

3.2.1. KPCA directions as normal vectors

For class c, we map the samples into a reproducing kernel Hilbert space (RKHS) $\mathcal{F}$ through a feature map $\phi : \mathbb{R}^m \to \mathcal{F}$ induced by the Gaussian kernel:

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{F}} = \exp\Big(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\Big). \tag{3.2}$$

The kernel matrix $\mathbf{K}_c \in \mathbb{R}^{N_c \times N_c}$ has entries $[\mathbf{K}_c]_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$ for $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}_c$. Let $\tilde{\mathbf{K}}_c$ be the centered kernel matrix,

$$\tilde{\mathbf{K}}_c = \mathbf{K}_c - \mathbf{1}_{N_c}\mathbf{K}_c - \mathbf{K}_c\mathbf{1}_{N_c} + \mathbf{1}_{N_c}\mathbf{K}_c\mathbf{1}_{N_c}, \quad \mathbf{1}_{N_c} = \tfrac{1}{N_c}\mathbf{1}\mathbf{1}^\top. \tag{3.3}$$

The eigendecomposition

$$\tilde{\mathbf{K}}_c\boldsymbol{\alpha}_{c,j} = \lambda_{c,j}\boldsymbol{\alpha}_{c,j}, \qquad \lambda_{c,1} \geq \lambda_{c,2} \geq \cdots \geq \lambda_{c,N_c} \geq 0 \tag{3.4}$$

yields the KPCA directions, with eigenvectors normalized as $\|\boldsymbol{\alpha}_{c,j}\|^2 = 1/\lambda_{c,j}$ for $\lambda_{c,j} > 0$. For any sample $\mathbf{x}$, its coordinate along the j-th KPCA direction is

$$f_{c,j}(\mathbf{x}) = \sum_{i=1}^{N_c} \alpha_{c,j}(i)\, \kappa(\mathbf{x}_i, \mathbf{x}). \tag{3.5}$$

We set the normal vector $\mathbf{w}_{c,j}$ of the j-th hyperplane to correspond to the j-th KPCA direction, represented implicitly through the kernel expansion [18]. The following two propositions establish why this choice is theoretically sound.

Proposition 3.1 (Variance decomposition in RKHS)**.** *Let class c consist of K_c sub-modalities with means $\boldsymbol{\psi}_{c,1}, \ldots, \boldsymbol{\psi}_{c,K_c} \in \mathcal{F}$ and within-submodality covariance operators $\boldsymbol{\Sigma}_{c,k}$. For any direction $\mathbf{v} \in \mathcal{F}$,*

$$\mathrm{Var}_{\mathrm{total}}(\mathbf{v}) = \sum_{k=1}^{K_c} \pi_{c,k} \left\|\mathbf{v}^\top(\boldsymbol{\psi}_{c,k} - \bar{\boldsymbol{\psi}}_c)\right\|^2 + \sum_{k=1}^{K_c} \pi_{c,k}\mathbf{v}^\top\boldsymbol{\Sigma}_{c,k}\mathbf{v}, \tag{3.6}$$

where $\pi_{c,k} = N_{c,k}/N_c$ and $\bar{\boldsymbol{\psi}}_c = \sum_k \pi_{c,k}\boldsymbol{\psi}_{c,k}$. Hence, the first KPCA direction $\mathbf{v}_1$, which maximizes $\mathrm{Var}_{\mathrm{total}}$, satisfies

$$\mathrm{Var}_{\mathrm{between}}(\mathbf{v}_1) \geq \mathrm{Var}_{\mathrm{total}}(\mathbf{v}_1) - \max_k \lambda_{\max}(\boldsymbol{\Sigma}_{c,k}). \tag{3.7}$$

Proof. The decomposition above is the law of total variance applied in $\mathcal{F}$. Since $\mathbf{v}_1$ attains the maximum of $\mathrm{Var}_{\mathrm{total}}$ and $\mathrm{Var}_{\mathrm{within}}(\mathbf{v}) \leq \max_k \lambda_{\max}(\boldsymbol{\Sigma}_{c,k})$ for every unit $\mathbf{v}$, the bound follows directly. □

Proposition 3.1 states that the first KPCA direction maximizes a lower bound on the between-submodality variance. When sub-modalities are well-separated (the between-submodality term dominates), $\mathbf{v}_1$ is near-optimal for discriminating them. The Gaussian kernel further promotes this behavior by approximately sphericizing within-submodality distributions in RKHS, as established in the kernel methods literature [18].

Proposition 3.2 (Optimality under isotropic covariance)**.** *Assume that within $\mathcal{F}$, the K_c sub-modalities of class c share identical, isotropic covariance operators, $\boldsymbol{\Sigma}_{c,k} = \sigma^2\mathbf{I}$ for all k, and that the sub-modalities have equal prior probabilities (or equivalently, equal misclassification costs). Then:*

1) The leading $K_c - 1$ KPCA directions span the range of the between-submodality scatter matrix.
2) For $K_c = 2$, the hyperplane with normal vector $\mathbf{v}_1$ and bias positioned at the midpoint of the two sub-modality means (projected onto $\mathbf{v}_1$) is the Bayes-optimal decision boundary under the 0–1 loss.

Proof. Under isotropy, $\mathrm{Var}_{\mathrm{within}}(\mathbf{v}) = \sigma^2$ is constant across all directions, so maximizing $\mathrm{Var}_{\mathrm{total}}$ coincides with maximizing $\mathrm{Var}_{\mathrm{between}}$. The eigenvectors of the total scatter therefore equal those of the between-submodality scatter. For $K_c = 2$, the between-submodality scatter is rank-1 with $\mathbf{v}_1 \propto \boldsymbol{\psi}_{c,1} - \boldsymbol{\psi}_{c,2}$. Under equal isotropic covariances and equal prior probabilities, the Bayes-optimal decision boundary under the 0–1 loss is the perpendicular bisector of the segment joining the two means. □

Remark. The isotropy assumption is idealized; it serves as a theoretical benchmark indicating when KPCA directions are provably optimal. In practice, the Gaussian kernel tends to make data distributions in RKHS more isotropic than in the input space. When sub-modalities overlap heavily, the first KPCA direction may not perfectly separate them—a limitation discussed in Section 5.

3.2.2. Hyperplane positioning

With the normal vector determined, the bias $b_{c,k}$ must place the hyperplane in the low-density gap between sub-modalities. We initialize $b_{c,k}^{(0)}$ at the median of the scalar projections $\{\mathbf{w}_{c,k}^\top \mathbf{x}_i\}_{\mathbf{x}_i \in \mathcal{X}_c}$, split the samples accordingly into $\mathcal{X}_{c,k}^+$ and $\mathcal{X}_{c,k}^-$, and refine:

$$b_{c,k} = -\mathbf{w}_{c,k}^\top \cdot \frac{\boldsymbol{\mu}_{c,k}^+ + \boldsymbol{\mu}_{c,k}^-}{2}, \qquad \boldsymbol{\mu}_{c,k}^\pm = \frac{1}{|\mathcal{X}_{c,k}^\pm|} \sum_{\mathbf{x}_i \in \mathcal{X}_{c,k}^\pm} \mathbf{x}_i. \tag{3.8}$$

This places $\mathcal{H}_{c,k}$ exactly midway between the two subclass centroids along the normal direction.

3.3. Hyperplane number selection via the eigenvalue elbow rule

A distinctive advantage of the hyperplane framework is that the number of hyperplanes can be determined via a principled, data-driven strategy. Existing subclass-based methods require the number of subclasses to be specified a priori or selected via grid search [16]. In GMLDA, the eigenvalue spectrum of $\tilde{\mathbf{K}}_c$—already computed when solving for normal vectors—directly guides this selection process.

The eigenvalues $\lambda_{c,1} \geq \lambda_{c,2} \geq \cdots \geq \lambda_{c,N_c}$ quantify the variance explained by successive KPCA directions. When class c contains K_c genuine sub-modalities, the first $K_c - 1$ eigenvalues are markedly larger than the remaining eigenvalues. These dominant eigenvalues account for cross-modality variation, while the subsequent smaller eigenvalues mainly describe intra-modality scatter and noise. This behavior creates a pronounced drop, referred to as the *elbow*, in the eigenvalue curve at index $K_c - 1$.

Definition 3.2 (Eigenvalue elbow rule)**.** *For a threshold $\tau > 1$, the estimated number of sub-modalities in class c is*

$$\hat{K}_c = \min\Big\{ j \in \{1, \ldots, N_c - 1\} \;\Big|\; \frac{\lambda_{c,j}}{\lambda_{c,j+1}} > \tau \Big\} + 1, \tag{3.9}$$

with $\hat{K}_c = 2$ if no j satisfies the condition. The number of hyperplanes is $\hat{K}_c - 1$.

The ratio $\lambda_{c,j}/\lambda_{c,j+1}$ measures the relative reduction in explained variance between consecutive KPCA directions. When this ratio exceeds τ, the $(j+1)$-th direction contributes less than $1/\tau$ of the variance captured by the j-th direction, which provides a natural criterion for identifying the elbow

point. The default value is set to $\tau = 2.0$, meaning the variance contribution is required to drop by at least half at the elbow position. This choice balances sensitivity and specificity: A smaller τ (e.g., 1.5) risks detecting spurious elbows due to sampling noise, while a larger τ (e.g., 3.0) may miss genuine but subtle modality structures. The value 2.0 is a widely adopted heuristic in covariance structure analysis and provides a conservative threshold that reliably identifies dominant spectral gaps. Sensitivity analysis (see supplementary material) confirms that GMLDA performance is stable for $\tau \in [1.5, 3.0]$.

When no j satisfies the ratio condition—indicating a smoothly decaying eigenvalue spectrum without a pronounced elbow—the data do not exhibit clear evidence of multiple sub-modalities. In this case, the fallback $\hat{K}_c = 2$ is adopted. This fallback is motivated by two considerations. First, setting $\hat{K}_c = 1$ would reduce the method to standard LDA for class c, forfeiting any potential gain from intra-class partitioning and potentially under-partitioning classes with weak but genuine multimodal structure. Second, using $\hat{K}_c = 2$ (one hyperplane) provides a minimal yet safe degree of intra-class subdivision: If the class is truly unimodal, the resulting two subclasses will largely overlap, and the scatter matrix structure will closely approximate that of standard LDA, causing little harm; if the class does contain latent sub-modalities, even a single hyperplane oriented along the dominant KPCA direction can capture a meaningful fraction of the between-submodality variation.

3.4. GMLDA: Hyperplane-partitioned LDA

Once each class has been partitioned, we obtain a collection of subclasses $\{\mathcal{X}_{c,k}\}$ with $\mathcal{X}_c = \bigcup_{k=1}^{K_c} \mathcal{X}_{c,k}$. LDA is then performed on these subclasses, treating each subclass as a separate entity. The scatter matrices are:

$$\mathbf{S}'_B = \sum_{c=1}^{C} \sum_{k=1}^{K_c} N_{c,k}(\boldsymbol{\mu}_{c,k} - \boldsymbol{\mu})(\boldsymbol{\mu}_{c,k} - \boldsymbol{\mu})^\top, \tag{3.10}$$

$$\mathbf{S}'_W = \sum_{c=1}^{C} \sum_{k=1}^{K_c} \sum_{\mathbf{x}_i \in \mathcal{X}_{c,k}} (\mathbf{x}_i - \boldsymbol{\mu}_{c,k})(\mathbf{x}_i - \boldsymbol{\mu}_{c,k})^\top, \tag{3.11}$$

where $\boldsymbol{\mu}_{c,k}$ and $N_{c,k}$ are the mean and size of the k-th subclass of class c and $\boldsymbol{\mu}$ is the global mean.

The hyperplane-partitioned scatter matrices are provably more discriminative than the original LDA scatter matrices:

Proposition 3.3 (Scatter matrix improvement)**.** *Let $\mathbf{S}_B, \mathbf{S}_W$ be the original scatter matrices (Eqs (1) and (2)). Then* $\operatorname{rank}(\mathbf{S}'_B) \geq \operatorname{rank}(\mathbf{S}_B)$ *and* $\operatorname{tr}(\mathbf{S}'_W) \leq \operatorname{tr}(\mathbf{S}_W)$*, with strict inequality whenever $K_c > 1$ for some class c , and the corresponding subclasses contain at least two nonempty subsets with distinct centroids.*

Proof. $\mathbf{S}'_B$ aggregates $\sum_c K_c \geq C$ subclass centroids, the span of which contains the C class centroids of $\mathbf{S}_B$, hence the rank cannot decrease. For the trace, the law of total variance decomposes the scatter of class c as $\sum_{\mathbf{x}_i \in \mathcal{X}_c} \|\mathbf{x}_i - \boldsymbol{\mu}_c\|^2 = \sum_k \sum_{\mathbf{x}_i \in \mathcal{X}_{c,k}} \|\mathbf{x}_i - \boldsymbol{\mu}_{c,k}\|^2 + \sum_k N_{c,k} \|\boldsymbol{\mu}_{c,k} - \boldsymbol{\mu}_c\|^2$. The first right-hand term is the class-c contribution to $\operatorname{tr}(\mathbf{S}'_W)$ and the left-hand term to $\operatorname{tr}(\mathbf{S}_W)$. The second right-hand term is strictly positive when $K_c > 1$, and the subclasses have distinct centroids. □

The projection matrix $\mathbf{W}$ is obtained via the Fisher criterion:

$$\mathbf{W}^* = \arg\max_{\mathbf{W}} \frac{|\mathbf{W}^\top \mathbf{S}'_B \mathbf{W}|}{|\mathbf{W}^\top \mathbf{S}'_W \mathbf{W}|}, \tag{3.12}$$

solved as the generalized eigenvalue problem $\mathbf{S}'_B\mathbf{W} = \lambda\mathbf{S}'_W\mathbf{W}$, retaining the top d eigenvectors. When the number of subclasses is large relative to the sample size, we regularize $\mathbf{S}'_W$ via $\tilde{\mathbf{S}}'_W = \mathbf{S}'_W + \epsilon\mathbf{I}$ with $\epsilon = 10^{-5} \cdot \mathrm{tr}(\mathbf{S}'_W)/m$ [19].

Algorithm 1 summarizes the complete procedure. We note two key properties: (i) KPCA is performed per class, consistent with the theory of Section 3.2; (ii) the number of hyperplanes $K_c - 1$ is determined automatically by the elbow rule (Definition 3.2), requiring no manual tuning.

Algorithm 1 GMLDA: Geometric partitioning-based multimodal LDA

Input: Dataset $\mathbf{X} \in \mathbb{R}^{m\times n}$, labels $\mathbf{y} \in \{1, \ldots, C\}^n$, kernel parameter σ, elbow threshold $\tau = 2.0$
Output: Projection matrix $\mathbf{W}$, low-dimensional embedding $\mathbf{Z}$
for each class $c = 1, \ldots, C$ **do**
    Construct class-specific kernel matrix $\mathbf{K}_c$ via Eq (3.2); center to obtain $\tilde{\mathbf{K}}_c$
    Eigendecomposition: $\tilde{\mathbf{K}}_c\boldsymbol{\alpha}_{c,j} = \lambda_{c,j}\boldsymbol{\alpha}_{c,j}$ (Eq (3.4))
    Determine K_c via the elbow rule (Eq (3.9))
    for $k = 1, \ldots, K_c - 1$ **do**
        Set normal vector direction via $\boldsymbol{\alpha}_{c,k}$ (Eq (3.5))
        Compute bias $b_{c,k}$ via Eq (3.8)
        Split the current subclass using $\mathcal{H}_{c,k} : \mathbf{w}_{c,k}^\top\mathbf{x} + b_{c,k} = 0$
    end for
end for
Assemble $\mathbf{S}'_B$, $\mathbf{S}'_W$ via Eqs. (3.10)–(3.11); regularize $\tilde{\mathbf{S}}'_W = \mathbf{S}'_W + \epsilon\mathbf{I}$
Solve $\tilde{\mathbf{S}}'^{-1}_W\mathbf{S}'_B\mathbf{W} = \lambda\mathbf{W}$; return top d eigenvectors as $\mathbf{W}$
return $\mathbf{W}$ and $\mathbf{Z} = \mathbf{W}^\top\mathbf{X}$

3.5. Computational complexity

The cost comprises three stages. (i) KPCA per class: $O(\sum_c(N_c^2D + N_c^3))$, dominated by the cubic eigendecomposition. When classes are balanced ($N_c \approx N/C$), this reduces to $O(N^3/C^2)$, a factor C^2 improvement over global KPCA. (ii) Hyperplane partitioning: $O(NDK)$ with $K = \sum_c(K_c - 1)$, which is linear in all parameters. (iii) LDA on subclasses: $O(ND^2 + D^3)$ for scatter matrix assembly and the generalized eigenvalue problem. The overall complexity is $O(\sum_c N_c^3 + ND^2)$. Unlike LFDA, NLDA, and subclass-based methods that require $O(N^2)$ pairwise distance computations or iterative neighborhood constructions, GMLDA's hyperplane partitioning stage is $O(NDK)$, and the KPCA cost is confined to individual classes, which can be processed in parallel.

4. Numerical experiments

In this section, the proposed GMLDA is compared with other LDA-type methods, including LDA [1] and its modified versions: RSLDA [2], TRLDA [3], ccLDA [4], LFDA [6], NLDA [7], and MPDA [8]. For all the aforementioned supervised dimensionality reduction methods, the KNN classifier is adopted to obtain the final classification accuracy of different methods. In addition, the support vector machine (SVM) with the radial basis function (RBF) kernel is also employed as a supplementary classifier; its results on all datasets are summarized in Table 6 for completeness. Hyperparameters of both classifiers are optimized via grid search with 5-fold cross-validation. All the codes involved in the above experiments are implemented in Python. The experimental datasets consist of a synthetic dataset, UCI benchmark datasets, handwritten digit datasets, facial image datasets, and object image datasets. To ensure statistical reliability, every experiment reported in Tables 2–6 is repeated 20 times with independent random training/test splits, and the average classification accuracy is reported. Standard deviations are reported alongside the mean values in the corresponding tables.

Parameter settings. The Gaussian kernel width σ in KPCA is set via the median heuristic: $\sigma = \text{median}\{\|\mathbf{x}_i - \mathbf{x}_j\| : \mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}_c\}$, computed per class. For the KNN classifier, the neighborhood size k is selected from $\{1, 3, 5, 7, 9\}$ via 5-fold cross-validation on the training set. For the RBF-SVM classifier, the regularization parameter C and kernel width γ are jointly tuned via grid search over $C \in \{10^{-3}, 10^{-2}, \ldots, 10^{3}\}$ and $\gamma \in \{10^{-3}, 10^{-2}, \ldots, 10^{3}\}$. The output dimensionality d for all methods is set to $C - 1$ (where C is the number of original classes), which is the standard setting for LDA-type methods and ensures a fair comparison. The elbow threshold is fixed at $\tau = 2.0$ throughout all experiments; the rationale for this default is discussed in Section 3.3.

4.1. Synthetic dataset

To intuitively demonstrate the idea of hyperplane segmentation for intra-class multi-modal data, a two-dimensional synthetic dataset is generated in this paper. As shown in Figure 1, the dataset consists of two classes (Class 0 and Class 1), and each class contains two data clusters following a Gaussian distribution. Each cluster is distributed at a different spatial position to simulate the intra-class multi-modal characteristics. To determine the reference direction for intra-class segmentation, KPCA is adopted to extract the first principal component direction of the samples in each class. This direction can characterize the main variation trend of the corresponding class data, so it is used as the normal vector of the local hyperplane for that class (as indicated by the arrows in the figure), where the blue arrow corresponds to Class 0 and the red arrow corresponds to Class 1. The dashed lines in the figure represent the intra-class hyperplanes of the two classes, respectively: The hyperplane of Class 0 is located in the upper region of the space, and the hyperplane of Class 1 is located in the lower region of the space, with the two hyperplanes being independent of each other and not interfering with one another. This figure clearly shows that for classes with an intra-class multi-modal structure, the normal vector obtained based on the KPCA method can effectively determine their local hyperplanes, which can accurately segment different sub-clusters within the class and lay a solid theoretical and data foundation for subsequent dimensionality reduction and classification tasks.

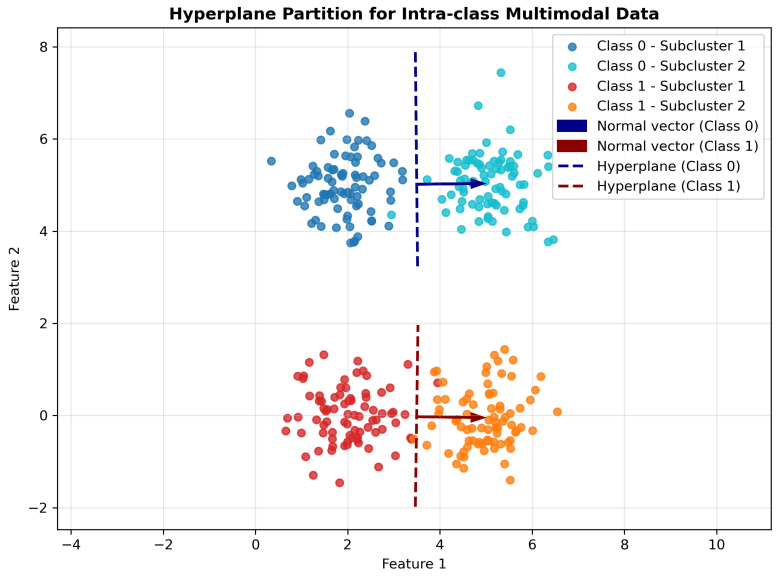

Figure 1. Hyperplane partition for intra-class multimodal data.

4.2. UCI datasets

In this subsection, the proposed method is tested on the datasets retrieved from the UCI machine learning repository, which are detailed in Table 1. First, a standardization process is performed on the raw datasets to transform them into a distribution with zero mean and unit variance, thereby eliminating the interference of dimensional differences among distinct features on subsequent analyses. Next, the standardized data of each class are partitioned into training and test subsets following a ratio of 7:3. Prior to conducting the experiments, data standardization is uniformly implemented, followed by dimensionality reduction via PCA with the constraint of preserving 95% of the cumulative explained variance to retain the primary information of the original data. Then the GMLDA method is carried out, and the corresponding KNN-based results are presented in Table 2. The SVM-based results are reported in Table 6. It can be observed that the proposed method achieves superior classification performance compared with other LDA-type methods on most of the datasets under both classifiers.

Table 1. Details of UCI datasets.

Dataset	Classes	Samples	Features
Wine	3	178	13
Ionosphere	2	351	34
Mushrooms	2	8124	22
Balance-scale	3	625	4
Iris	3	150	4
Wave form	3	5000	21
Wholesale customers	2	440	6
Yeast	10	1484	8
Banana	2	5312	2
Thyroid	3	7200	21

Table 2. Experimental results on UCI datasets (mean ± std. over 20 runs, %).

	LDA	RSLDA	TRLDA	ccLDA	LFDA	NLDA	MPDA	GMLDA
Wine	96.30±1.53	92.59±1.24	62.50±1.25	97.22±0.99	81.48±1.67	94.44±1.88	95.51±1.79	**100.00**±0.00
Ionosphere	85.85±1.24	85.85±1.79	77.91±1.58	78.87±1.89	86.79±1.73	83.69±1.22	89.62±0.00	**90.57**±1.27
Mushrooms	99.96±0.00	89.38±1.23	89.78±1.87	85.97±1.45	60.25±1.58	99.97±1.39	100.0±0.00	**100.0**±0.00
Balance-scale	**92.00**±1.23	83.51±1.99	45.09±1.98	87.20±1.78	87.23±1.76	78.72±1.65	91.49±1.26	89.89±1.33
Iris	95.56±1.20	51.11±1.89	95.56±1.47	**96.67**±1.78	95.56±1.24	91.11±1.29	91.11±1.56	95.56±1.25
Wave form	85.13±1.29	85.73±1.00	85.76±0.82	83.90±1.29	32.33±1.37	85.60±1.46	81.45±1.78	**86.13**±1.47
Yeast	57.62±1.38	54.93±1.89	56.89±1.47	58.59±1.78	39.91±1.47	56.73±1.38	53.14±1.48	**62.33**±1.58
Banana	71.01±0.99	88.18±0.94	67.26±0.29	68.21±1.11	87.74±1.38	62.58±1.48	88.24±1.85	**89.25**±1.59
Thyroid	95.38±1.47	94.57±1.38	92.68±1.22	87.04±0.98	92.00±0.79	95.38±0.92	95.38±0.79	**96.92**±1.44

4.3. Handwritten digit datasets

In this subsection, the proposed method is evaluated on the handwritten digit datasets USPS and MNIST. The USPS dataset consists of handwritten digit images extracted from real mails, with a total of approximately 9,298 samples (including 7,291 training samples and 2,007 test samples). Each image is a grayscale map with a resolution of 16×16 pixels. Some typical images of the USPS dataset are illustrated in Figure 2(a). The MNIST dataset contains 70,000 samples (including 60,000 training samples and 10,000 test samples). Each image is a grayscale map of 28×28 pixels, where the background is black and the digits are white. The samples have undergone preprocessing steps such as centering and size normalization, resulting in low noise levels. Each category is composed of five subclasses. Similar preprocessing operations to those applied to the aforementioned UCI benchmark datasets are also implemented here. In each dataset, we attempt to identify whether a digit is odd or even. Thus, the task of handwritten digit recognition is transformed into a binary classification problem. Experimental results demonstrate that the classification accuracy rates on the USPS and MNIST datasets reach 96.45% and 93.17%, respectively. These results demonstrate that the proposed GMLDA method outperforms typical methods of multi-modal LDA (e.g., LFDA, NLDA, MPDA) on handwritten datasets. The KNN-based results are shown in Table 3, and the SVM-based results are summarized in Table 6.

4.4. Face image datasets

In this subsection, our proposed method is evaluated on three face image datasets, namely the Extended Yale B dataset (see Figure 2(b)), ORL Face dataset, and Yale dataset. Similar preprocessing procedures to those adopted for the aforementioned UCI benchmark datasets are carried out for these face datasets. Numerical experiments of the GMLDA method are conducted on each of the above datasets, with the KNN-based results presented in Table 4 and the SVM-based results summarized in Table 6. The proposed method achieves superior classification performance on all three datasets under both classifiers. Specifically, the classification accuracy of our method reaches 97.5%, 99.17%, and 98% on the Extended Yale B, ORL Face, and Yale datasets, respectively.

4.5. Object image datasets

In this subsection, our proposed method is evaluated on object image datasets including the COIL-20 dataset (see Figure 2(c)) and the COIL-100 dataset. The KNN-based results are reported in Table 5, and the SVM-based results are summarized in Table 6, both confirming that the proposed GMLDA method performs better than other tested LDA-type methods on object image datasets.

(a)

(b)

(c)

Figure 2. Typical images of datasets: (a) USPS, (b) Extended Yale B, (c) COIL-20.

Table 3. Experimental results on handwritten digit datasets (mean ± std. over 20 runs, %).

	LDA	RSLDA	TRLDA	ccLDA	LFDA	NLDA	MPDA	GMLDA
USPS	93.69±1.29	91.18±1.92	93.33±1.19	82.83±1.29	89.86±1.76	94.00±1.79	57.35±1.90	**96.52**±1.80
MNIST	86.55±1.02	91.00±0.98	88.25±0.87	85.00±1.87	70.30±1.02	89.13±1.02	89.93±1.04	**93.07**±1.23

Table 4. Experimental results on face image datasets (mean ± std. over 20 runs, %).

	LDA	RSLDA	TRLDA	ccLDA	LFDA	NLDA	MPDA	GMLDA
Extended Yale B	97.30±1.07	77.23±0.99	96.96±0.97	86.37±0.43	95.47±1.90	97.04±1.02	96.67±1.89	**97.50**±1.79
ORL face	95.55±1.20	90.65±1.29	90.83±1.26	89.29±1.48	85.00±1.48	98.33±1.59	99.17±1.49	**99.17**±1.39
Yale	97.00±1.48	54.00±1.58	70.00±1.59	93.33±1.48	64.00±1.09	84.00±1.99	98.00±1.38	**98.00**±1.75

Table 5. Experimental results on object image datasets (mean ± std. over 20 runs, %).

	LDA	RSLDA	TRLDA	ccLDA	LFDA	NLDA	MPDA	GMLDA
COIL-20	85.42±1.22	80.32±1.90	91.13±0.99	96.67±0.58	89.58±0.50	98.38±1.88	84.49±1.22	**99.07**±0.95
COIL-100	91.81±1.29	85.46±1.48	96.81±1.55	95.00±1.33	88.56±1.39	94.31±1.50	96.81±1.48	**96.85**±1.38

Table 6. SVM-based classification accuracy on all datasets (mean ± std. over 20 runs, %).

	LDA	RSLDA	TRLDA	ccLDA	LFDA	NLDA	MPDA	GMLDA
Wine	96.30±1.99	92.59±1.33	64.81±1.22	98.15±1.33	83.33±1.34	96.30±1.89	96.30±1.09	**100.00**±0.00
Ionosphere	85.85±1.88	84.91±1.77	79.25±1.88	79.25±1.76	87.74±1.22	85.85±0.78	89.62±1.44	**91.51**±1.66
Mushrooms	99.96±0.00	89.87±1.99	89.92±1.64	87.03±1.76	62.14±1.46	99.97±0.00	100.00±0.00	**100.00**±0.00
Balance-scale	**92.55**±1.33	83.51±1.76	47.34±1.76	88.30±1.59	88.83±1.78	81.38±1.33	91.49±1.58	90.43±1.68
Iris	95.56±1.48	53.33±1.69	95.56±1.78	**97.78**±1.06	95.56±1.77	93.33±1.78	93.33±1.47	95.56±1.49
Waveform	85.27±1.78	85.80±1.45	85.93±1.57	84.13±1.48	33.27±3.49	85.87±2.22	82.00±1.48	**86.60**±2.33
Yeast	58.07±3.21	55.61±3.49	57.40±3.44	59.19±2.44	41.26±2.48	57.17±2.08	54.04±2.98	**62.78**±2.49
Banana	72.15±2.34	88.55±2.94	67.82±2.00	68.53±1.39	88.02±2.39	63.47±2.30	88.62±2.99	**89.51**±3.10
Thyroid	95.38±1.39	94.62±1.39	92.77±1.29	87.23±1.39	92.15±1.49	95.48±1.39	95.50±1.29	**96.92**±2.38
USPS	93.82±1.33	92.05±1.39	93.67±1.22	84.15±2.33	90.12±2.55	94.28±3.22	59.48±3.22	**96.84**±1.33
MNIST	87.03±1.28	22.50±1.45	88.67±1.56	85.43±2.33	71.23±3.22	89.57±2.33	90.23±2.66	**93.43**±2.55
Extended Yale B	97.64±2.33	78.45±3.33	97.16±2.92	86.85±2.19	96.08±1.55	97.30±1.22	96.89±0.22	**97.84**±0.38
ORL face	98.33±0.24	91.67±0.98	91.67±1.23	90.00±2.33	85.83±3.33	98.33±0.77	99.17±0.00	**99.17**±0.00
Yale	**98.10**±0.12	55.24±4.33	71.43±2.34	93.33±2.44	65.71±2.34	85.71±2.44	98.00±1.00	98.00±0.58
COIL-20	85.42±3.33	80.32±2.34	91.67±1.34	96.67±1.34	90.28±1.57	98.61±1.48	85.42±1.48	**99.54**±0.00
COIL-100	92.08±1.22	85.74±1.34	96.90±1.22	95.23±1.28	88.89±1.29	94.54±1.29	97.04±1.29	**97.08**±0.99

4.6. Ablation study

To quantitatively dissect the independent contribution of each component in the proposed GMLDA method, and clearly verify that hyperplane geometric partitioning acts as the core innovation while

KPCA only serves as an auxiliary module for solving segmentation normals, we design five ablation variants for controlled comparison. The detailed settings of each variant are listed as follows:

1) GMLDA: The complete proposed method. Class-wise KPCA is adopted to calculate hyperplane normals, and the eigenvalue elbow criterion is utilized to adaptively determine the number of hyperplanes. All experimental configurations remain consistent across all compared schemes.
2) PCA-hyperplane: Retain the whole hyperplane partitioning framework, replace KPCA with linear PCA to generate hyperplane normal vectors, and keep other parameters unchanged.
3) Random-hyperplane: Maintain the core hyperplane segmentation mechanism, and random unit vectors are directly taken as hyperplane normals without any feature decomposition.
4) K-means-LDA: Abandon hyperplane geometric partitioning. Instead, K-means clustering with the identical number of subclasses as GMLDA is used to split intra-class samples, followed by standard LDA.
5) LDA: Baseline method without any intra-class subdivision. Each category is represented by a single global centroid for discriminant projection.

Table 7. Ablation results of different model components.

Dataset	GMLDA	PCA-Hyperplane	Random-Hyperplane	K-means-LDA	LDA
Balance-scale	89.89	88.83	90.43	83.19	89.36
Iris	95.56	93.33	91.11	91.11	88.89
Wine	100.00	100.00	98.15	100.00	94.44
Ionosphere	93.57	90.57	83.02	93.40	90.57
Mushrooms	100.00	100.00	100.00	99.71	96.55
Waveform	86.13	84.47	85.33	84.80	85.60
Yeast	62.33	61.43	60.54	58.74	60.99
Banana	89.25	89.18	89.18	89.18	89.25
Thyroid	96.92	96.92	96.92	96.92	93.85
USPS	96.52	92.55	91.56	92.52	90.45
MNIST	93.07	89.30	89.77	88.07	90.87
Extended Yale B	97.50	97.33	97.37	97.29	97.30
ORL face	99.17	95.33	92.34	94.50	95.55
Yale	98.0	94.00	98.00	96.00	97.00
COIL-20	99.07	95.31	98.38	93.61	85.42
COIL-100	96.85	96.30	96.11	96.11	91.81

For fair evaluation, all datasets are split into training and test subsets at a ratio of 7:3, and each experiment is repeated 20 times to record the average classification accuracy (in percentage). The comprehensive ablation results are summarized in Table 7. As observed from the table, all hyperplane-based variants outperform standard LDA and K-means-LDA on most multimodal datasets, which verifies that geometric space partitioning is the core contributor to performance improvement; full GMLDA achieves mild accuracy gains over PCA-hyperplane and random-hyperplane, proving KPCA only serves as an auxiliary tool to optimize segmentation directions rather than the core framework. For datasets with weak intra-modal differences, all methods deliver close results since subdivision brings

limited promotion to discriminative capability. Overall, the ablation experiments fully validate our core idea that hyperplane geometric partitioning dominates the performance of the proposed GMLDA, while KPCA plays a secondary supporting role.

5. Conclusions

In this paper, we propose an improved multimodal linear discriminant analysis method GMLDA based on geometric partitioning. GMLDA splits intra-class multimodal data using hyperplanes, and takes kernel principal component directions as hyperplane normal vectors to capture core data distribution patterns. Jointly optimized hyperplanes tighten intra-class submodality distributions and expand inter-class margins, boosting feature discrimination. Experimental results on a range of benchmark datasets demonstrate that GMLDA achieves competitive or superior performance against existing multimodal LDA algorithms. This work also has two clear limitations. First, per-class KPCA eigendecomposition brings $O(N_c^3)$ computational complexity. Though calculated class-wisely, it becomes a bottleneck for extremely large classes; scalable approximations like the Nyström method can be studied for future optimization. Second, GMLDA adopts LDA's linear projection framework, which lacks expressive power for data with highly nonlinear class boundaries. Kernelized or deep extensions of our framework are worthy follow-up research directions.

Author contributions

Xinyi Han: Conceptualization, Methodology, Software, Investigation, Writing-original draft, Funding acquisition; Chaojie Wang: Supervision, Validation, Writing-review & editing. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) in the creation of this article.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 12001022) and the Disciplinary funding of Beijing Technology and Business University (No. STKY202508).

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. R. A. Fisher, The use of multiple measurements in taxonomic problems, *Annals of Eugenics*, **7** (1936), 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

2. J. Wen, X. Z. Fang, J. R. Cui, L. K. Fei, K. Yan, Y. Chen, et al., Robust sparse linear discriminant analysis, *IEEE T. Circ. Syst. Vid.*, **29** (2019), 390–403. https://doi.org/10.1109/TCSVT.2018.2799214
3. J. Y. Wang, L. Wang, F. P. Nie, X. L. Li, A novel formulation of trace ratio linear discriminant analysis, *IEEE T. Neur. Net. Lear.*, **33** (2022), 5568–5578. https://doi.org/10.1109/tnnls.2021.3071030
4. Y. W. Pang, S. Wang, Y. Yuan, Learning regularized LDA by clustering, *IEEE T. Neur. Net. Lear.*, **25** (2014), 2191–2201. https://doi.org/10.1109/tnnls.2014.2306844
5. S. Boutemedjet, N. Bouguila, D. Ziou, A hybrid feature extraction selection approach for high-dimensional non-Gaussian data clustering, *IEEE T. Pattern Anal.*, **31** (2009), 1429–1443. https://doi.org/10.1109/tpami.2008.155
6. M. Sugiyama, Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis, *J. Mach. Learn. Res.*, **8** (2007), 1027–1061.
7. F. Zhu, J. B. Gao, J. Yang, N. Ye, Neighborhood linear discriminant analysis, *Pattern Recogn.*, **123** (2022), 108422. https://doi.org/10.1016/j.patcog.2021.108422
8. Y. Zhou, S. L. Sun, Manifold partition discriminant analysis, *IEEE T. Cybernetics*, **47** (2017), 830–840. https://doi.org/10.1109/tcyb.2016.2529299
9. T. Hastie, R. Tibshirani, Discriminant analysis by Gaussian mixtures, *J. R. Stat. Soc. B*, **58** (1996), 155–176. https://doi.org/10.1111/j.2517-6161.1996.tb02073.x
10. M. L. Zhu, A. M. Martinez, Subclass discriminant analysis, *IEEE T. Pattern Anal.*, **28** (2006), 1274–1286. https://doi.org/10.1109/tpami.2006.172
11. N. Gkalelis, V. Mezaris, I. Kompatsiaris, T. Stathaki, Mixture subclass discriminant analysis link to restricted gaussian model and other generalizations, *IEEE T. Neur. Net. Lear.*, **24** (2013), 8–21. https://doi.org/10.1109/tnnls.2012.2216545
12. Y. T. Tao, J. Yang, H. Y. Chang, Enhanced iterative projection for subclass discriminant analysis under EM-alike framework, *Pattern Recogn.*, **47** (2014), 1113–1125. https://doi.org/10.1016/j.patcog.2013.07.001
13. K. Chumachenko, J. Raitoharju, A. Iosifidis, M. Gabbouj, Speed-up and multi-view extensions to subclass discriminant analysis, *Pattern Recogn.*, **111** (2021), 107660. https://doi.org/10.1016/j.patcog.2020.107660
14. H. F. Zhang, K. Liu, Y. H. Zhang, J. H. Lin, TRANS-CNN-based gesture recognition for mmWave radar, *Sensors*, **24** (2024), 1800. https://doi.org/10.3390/s24061800
15. R. X. Lu, Y. J. Cai, J. Y. Zhu, F. P. Nie, H. Yang, Dimension reduction of multimodal data by auto-weighted local discriminant analysis, *Neurocomputing*, **461** (2021), 27–40. https://doi.org/10.1016/j.neucom.2021.06.035
16. H. Wan, H. Wang, G. D. Guo, X. Wei, Separability-oriented subclass discriminant analysis, *IEEE T. Pattern Anal.*, **40** (2018), 409–422. https://doi.org/10.1109/tpami.2017.2672557
17. Y. Zeng, Z. D. Jia, W. Liang, S. F. Gu, Fault diagnosis based on variable-weighted separability-oriented subclass discriminant analysis, *Comput. Chem. Eng.*, **129** (2019), 106514. https://doi.org/10.1016/j.compchemeng.2019.106514

18. B. Schölkopf, A. Smola, K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, *Neural Comput.*, **10** (1998), 1299–1319. https://doi.org/10.1162/089976698300017467

19. J. H. Friedman, Regularized discriminant analysis, *J. Am. Stat. Assoc.*, **84** (1989), 165–175. https://doi.org/10.1080/01621459.1989.10478752

AIMS Press