基于投影的多组二次判别分析

Yuchao Wang 统计与数据科学系，清华大学；
Tianying Wang*统计系，科罗拉多州立大学

arXiv:2601.05415v1 [stat.ME] 8 Jan 2026

摘要

多组分类问题广泛存在于诸多预测与决策场景中，包括流行病学、基因组学、金融和图像识别等应用。尽管分类方法已取得长足发展，但现有文献大多聚焦于二分类问题，而已有的扩展方法通常难以为多组场景提供足够的灵活性。近期研究已将线性判别分析扩展到多组情形，但为了处理非线性决策边界和组特异性协方差模式等复杂结构，仍需要更通用的方法。

我们提出多组二次判别分析（MGQDA），这是一种基于二次判别分析的多组分类方法。MGQDA 将高维预测变量投影到低维子空间中，从而在实现准确分类的同时，捕捉组特异性协方差结构中的非线性和异质性。我们推导了包括变量选择一致性在内的理论保证，以支持该方法的可靠性。在模拟研究和一项基因表达应用中，与现有方法相比，MGQDA 达到了具有竞争力或更优的预测性能，同时能够选择组特异性的信息变量，表明其在高维多组分类问题中具有实际应用价值。本文的补充材料可在线获取。

关键词：高维统计学、变量选择、组惩罚、块坐标下降

*通讯作者：Tianying.Wang@colostate.edu

1 引言

判别分析广泛应用于众多领域的分类任务，包括模式识别、《生物信息学》、金融和流行病学（Zhao 等人，1998；Lˆe Cao 等人，2011；Blasco 等人，2015；Chen 等人，2020）。线性判别分析（线性判别分析）因其简单且计算高效，长期以来备受青睐，尤其适用于低维情形。收缩估计量（Friedman，1989；Xu 等人，2009）和惩罚判别分析（Witten 和 Tibshirani，2011）等方法学进展提升了其稳定性，并拓展了其在高维数据中的适用范围。然而，线性判别分析假设各组具有相同的协方差矩阵，这限制了它在存在组特异性变异时的有效性。

二次判别分析（QDA）通过允许不同组具有各自的协方差矩阵，放宽了这一假设，因此在处理异质数据时更加灵活。在高维情形下，近期研究提出了惩罚似然方法（Danaher 等人，2014；Cai 和 Zhang，2021）和变量选择框架（Gaynanova 和 Wang，2019；Wu 和 Hao，2022），但这些方法主要针对两组问题。将二次判别分析扩展到多组分类会带来额外挑战，包括需要估计的参数数量增加、可扩展性问题，以及高维情况下可解释性降低。这些困难限制了二次判别分析在多组情形中的广泛应用，因此许多现有的多组分类方法仍依赖等协方差假设。

为应对这些挑战，我们提出多组二次判别分析（MGQDA），这是一个将二次判别规则的灵活性与内置变量选择相结合的框架。MGQDA 专为多组分类而设计，能够适应复杂的数据结构，且不要求各组共享同一个协方差矩阵。通过为所有组构建联合投影向量，MGQDA 将观测投影到低维子空间中，避免直接估计高维协方差矩阵或精度矩阵，从而实现计算高效的处理流程。

我们证明了 MGQDA 的变量选择一致性，表明它在适当条件下能够识别所有相关变量。在模拟和真实数据分析中，与现有方法相比，MGQDA 的分类准确率具有竞争力，甚至更高，同时选出的预测变量集规模更小，也更易于解释，因此为高维多组分类提供了一种可扩展且实用的工具。

本文其余部分安排如下。第 2 节介绍 MGQDA

框架和稀疏估计程序。第 3 节给出高维情形下变量选择一致性的理论结果。第 4 节和第 5 节通过模拟和真实数据应用报告实证评估。最后，第 6 节讨论 MGQDA 的更广泛影响，并概述可能的扩展。

在继续之前，我们先介绍本文中使用的符号。对于向量$v \in \mathbb{R}^p$，定义 $\|v\|_2 = \sqrt{\sum_{i=1}^p v_i^2}$、 $\|v\|_1 = \sum_{i=1}^p |v_i|$和 $\|v\|_\infty = \max_i |v_i|$. 。令 e_j 表示一个单位向量，其第 j个分量等于 1。对于一个 $p \times n$ 矩阵 M 以及子集 $S \subset \{1, \ldots, p\}$和 $T \subset \{1, \ldots, n\}$，其中 $\text{card}(S) = s \leq p$ 且 $\text{card}(T) = t \leq n$，令 m_i 表示 M的第 i行， M_S 表示由索引属于 S的行组成的 $s \times n$ 子矩阵， M_{ST} 表示由行索引属于 S 且列索引属于 T的 $s \times t$子矩阵。矩阵范数定义为$|||M|||_{\infty,2} = \max_i \|m_i\|_2$、 $|||M|||_\infty = \max_i \|m_i\|_1$、 $|||M|||_2 = \sup_{\|v\|_2=1} \|Mv\|_2$以及$\|M\|_F = \sqrt{\sum_i \sum_j m_{ij}^2}$. 。对于序列 a_n 和 b_n，若存在某个常数 $C > 0$ 使得 $a_n \leq C b_n$，则记为 $a_n = \mathcal{O}(b_n)$；若当 $n \to \infty$时 $a_n / b_n \to 0$，则记为 $a_n = o(b_n)$。

2 方法论

考虑从随机对（X, Y）中抽取的 n 个独立观测值（$X_1, Y_1 \ldots,$（X_n, Y_n））， 其取值属于 $\mathbb{R}^p \times \{1, \ldots, G\}$， 其中 G 表示组的数量。对于每个$g \in \{1, \ldots, G\}$， 令 $\Sigma_g = \text{cov}(X \mid Y = g)$ 表示组内协方差矩阵。组间总体协方差矩阵定义为 $\Sigma_B = \sum_{g=1}^G \pi_g (\mu_g - \mu)(\mu_g - \mu)^\top$,其中 $\pi_g = P(Y = g)$ 为组特定概率， $\mu_g = E(X \mid Y = g)$ 为组特定均值， $\mu = \sum_{g=1}^G \pi_g \mu_g$ 为总体均值。

2.1 背景与动机

我们简要回顾等协方差假设下的费舍尔判别分析$\Sigma_1 = \cdots = \Sigma_G = \Sigma$。令 $\Sigma_B = \sum_{g=1}^G \pi_g (\mu_g - \mu)(\mu_g - \mu)^\top$ 表示组间协方差。费舍尔方法寻求能够最大化瑞利商 $\phi^\top \Sigma_B \phi / \phi^\top \Sigma \phi$ 的判别方向 $\phi_1, \ldots, \phi_{G-1}$，并要求后续方向与之前获得的方向满足 Σ–正交:

$$\phi_1 = \arg\max_{\phi \neq 0} \frac{\phi^\top \Sigma_B \phi}{\phi^\top \Sigma \phi}, \phi_j = \arg\max_{\phi \neq 0} \frac{\phi^\top \Sigma_B \phi}{\phi^\top \Sigma \phi} \text{ subject to } \phi^\top \Sigma \phi_k = 0,\ k < j,\ j = 2, \ldots, G-1.$$

这一序列等价于广义特征值问题 $\Sigma_B v = \lambda \Sigma v$。令 $\Phi = (\phi_1, \ldots\ldots, \phi_{G-1})$ 表示由对应于最大的非零广义特征值的特征向量构成的投影基（最多有 $G - 1$ 个）。子空间 $\text{col}(\Phi)$

是唯一的；该子空间的任意基在 Σ–正交变换下（当相关特征值为单重时，还可相差一个缩放因子或符号）均可确定（Rao，1948；Golub 和 Van Loan，2013）。对于新的观测 $X_{\text{new}} \in \mathbb{R}^p$，总体分类规则 h_Φ 为

$$h_\Phi(X_{\text{new}}) = \arg\min_{1\le g\le G} (X_{\text{new}} - \mu_g)^\top \Phi(\Phi^\top \Sigma \Phi)^{-1}\Phi^\top (X_{\text{new}} - \mu_g) \; - \; 2\log \pi_g. \qquad (1)$$

因此，Fisher 判别法确定的子空间维数至多为 $G-1$，随后可在投影空间中使用公式 (1) 进行分类。确定方向的步骤仅依赖二阶矩，不要求正态性；而在具有共同协方差矩阵的多元正态类别下，公式 (1) 与贝叶斯规则一致。

等协方差假设在实践中往往不切实际。二次判别分析（QDA）允许各类别具有各自的协方差矩阵，从而放宽了这一假设。对于新的观测值 X_{new}，其总体判别规则为

$$h_{QDA}(X_{\text{new}}) = \arg\min_{1\le g\le G} (X_{\text{new}} - \mu_g)^\top \Sigma_g^{-1}(X_{\text{new}} - \mu_g) + \log|\Sigma_g| \; - \; 2\log \pi_g. \qquad (2)$$

2.2 所提方法

我们提出一种方法，构造适用于各组协方差结构的投影向量。该方法将数据投影到低维空间，再应用判别规则，从而结合了二次判别分析的灵活性，以及费舍尔方法的降维能力和可解释性。

对于每个组 $g = 1, \ldots, G$，令 $\bar{X}_g$ 表示样本均值，$\hat{\Sigma}_g$ 表示样本协方差矩阵。样本组间协方差定义为 $\hat{\Sigma}_B = \sum_{g=1}^{G} \binom{n_g}{n} (\bar{X}_g - \bar{X})(\bar{X}_g - \bar{X})^\top$，其中 $\bar{X}$ 为总体样本均值，n_g 为第 g 组中的观测值数量。仿照公式 (1) 和 (2) 中的总体判别规则，我们将 μ_g 替换为 $\bar{X}_g$，将 Σ_g 替换为 $\hat{\Sigma}_g$，并利用样本组间协方差 Σ_B 构造各组特定的投影方向。

对于每个 $g = 1, \ldots, G$，我们将 Φ_g 定义为矩阵束 $(\hat{\Sigma}_B, \hat{\Sigma}_g)$ 的非零广义特征向量矩阵，即 $\hat{\Sigma}_B v = \lambda \hat{\Sigma}_g v$. 该定义无需构造 $\hat{\Sigma}_g^{-1}$。当 $n_g > p$ 且 $\hat{\Sigma}_g \succ 0$ 时，问题可化为对 $\hat{\Sigma}_g^{-1}\hat{\Sigma}_B$ 进行熟知的特征分解。在满足 $p \ge n_g$ 的高维情形下，我们通过式（7）—（8）中的凸规划估计投影，从而得到 $\hat{\Omega}$。对于分类，我们设 $A_g(\hat{B}) = \hat{B}^\top \hat{\Sigma}_g \hat{B}$，其中 $\hat{B} \in \{\hat{\Phi}, \hat{\Omega}\}$。如果 $A_g(\hat{B})$ 是奇异的，则使用 Moore–Penrose 伪逆 $A_g(B)^+$ 以及伪行列式 $\widehat{\text{pdet}}\{A_g(B)\}$；参见附录 S.3.2。

因此，我们定义 $\widehat{\Phi} = [\widehat{\Phi}_1, \ldots, \widehat{\Phi}_G]$，其中 $\widehat{\Phi} \in \mathbb{R}^{p\times G(G-1)}$ 也是一个投影基；对于新的观测 $X_{\text{new}} \in \mathbb{R}^p$，分类规则为

$$\hat{h}_{\widehat{\Phi}}(X_{\text{new}}) = \arg\min_{1\le g\le G} (X_{\text{new}} - \bar{X}_g)^\top \widehat{\Phi}(\widehat{\Phi}^\top \widehat{\Sigma}_g \widehat{\Phi})^{-1} \widehat{\Phi}^\top (X_{\text{new}} - \bar{X}_g) + \log\left|\widehat{\Phi}^\top \widehat{\Sigma}_g \widehat{\Phi}\right| - 2\log(n_g/n). \tag{3}$$

这里，$\widehat{\Phi}$ 是总体基 Φ 的样本对应量。它将各组的组别基 $\{\widehat{\Phi}_g\}_{g=1}^G$ 按列拼接而成，其列张成了跨越所有组的联合判别子空间。公式（3）等价于对投影特征 $\widehat{\Phi}^\top X$应用二次判别规则。

我们的投影基不同于公式（1）中的投影基，因为我们分别纳入了各组的协方差结构，而不是在假设协方差相同的情况下汇总数据。这样既放宽了等协方差条件，又保留了 Fisher 判别规则中的投影思想，从而得到一种灵活且易于解释的分类器。然而，在高维情形下，估计 Φ_g 颇具挑战，因为当 p 与 n_g相当或更大时，代入估计量 Σ_B和 Σ_g 会变成奇异矩阵。因此，我们将在下一节介绍一种稀疏估计方法及相应的优化算法。

2.3 稀疏估计

在高维情形下，直接通过每个组 g 上对应于 $\widehat{\Sigma}_g^{-1}\widehat{\Sigma}_B$ 的非零特征向量来估计 Φ 颇具挑战，因为当 p 相对于 n 较大时，这种估计缺乏准确性。若没有额外的结构，式 (3) 中的样本分类规则在 $p \gg n$ 时可能表现不佳（Gaynanova 等人，2016）。

为了解决这一问题，我们施加稀疏性约束，并构造一个凸优化问题来估计 Φ。直接以 $\widehat{\Sigma}_g^{-1}\widehat{\Sigma}_B$ 的特征向量为基础构造目标函数较为困难，因此我们改为基于目标投影基的分块表示构造损失函数。

命题 2.1. 我们有 $\Sigma_B = \Gamma\Gamma^\top$ 和 $\widehat{\Sigma}_B = \widehat{\Gamma}\widehat{\Gamma}^\top$，其中 $\Gamma \in \mathbb{R}^{p\times(G-1)}$。此外，$\Gamma$ 的第 r 列具有如下形式：$\Gamma_r = \left[\sqrt{\pi_{r+1}}\{\sum_{i=1}^r \pi_i(\mu_i - \mu_{i+1})\}\right] / \sqrt{\sum_{i=1}^r \pi_i \sum_{i=1}^{r+1} \pi_i}$；而 $\widehat{\Gamma}$ 的第 r 列具有如下形式：$\widehat{\Gamma}_r = \left[\sqrt{n_{r+1}}\{\sum_{i=1}^r n_i(\bar{X}_i - \bar{X}_{r+1})\}\right] / \sqrt{n\sum_{i=1}^r n_i \sum_{i=1}^{r+1} n_i}$，其中 $r = 1, \ldots, G-1$。

该命题的证明与 Gaynanova 等人 (2016) 中的命题 2 类似。现在我们引入总体替代量 $\Theta_g := (\Sigma_g + \Sigma_B)^{-1}\Gamma$, $g = 1$，……，G，作为 Φ_g 的替代方案。其理论依据是，Θ_g 和 Φ_g 张成相同的判别

子空间，因此在投影 QDA 规则下会得出相同的分类决策。具体而言，我们在命题 2.2–2.3 中形式化了 Θ_g 与 Φ_g 之间的等价性，以及投影 QDA 规则的基不变性。

命题 2.2。对于每个组 g，Θ_g 和 Φ_g 的列空间相同：$\mathrm{col}(\Theta_g) = \mathrm{col}(\Phi_g)$，且存在一个可逆矩阵 $R_g \in \mathbb{R}^{(G-1)\times(G-1)}$，使得 $\Theta_g = \Phi_g R_g$。

命题 2.3。设 B 是张成某个判别子空间的任意满秩基，并定义投影 *QDA* 规则 $\delta_g(x;\ B) = (x-\mu_g)^\top B\{B^\top \Sigma_g B\}^{-1} B^\top (x-\mu_g) + \log\det(B^\top \Sigma_g B) - 2\log\,\pi_g$. 那么，对于任意可逆矩阵 R，分类器保持不变：$\arg\min_g \delta_g(x;\ BR) = \arg\min_g \delta_g(x;\ B)$。

命题 2.2–2.3 的证明见附录第 S.3.1 节。当$A_g(B) = B^\top \Sigma_g B$ 为奇异矩阵时（例如 $p \gg n_g$ 时经常如此），命题 2.3 中的逆矩阵和行列式分别替换为 Moore–Penrose 伪逆和伪行列式：$\delta_g(x;\ B) := (x-\mu_g)^\top B\{A_g(B)\}^{+} B^\top (x-\mu_g) + \log \mathrm{pdet}\{A_g(B)\} - 2\log\,\pi_g$. 在这些广义矩阵运算下，不变性结论仍然成立；详细讨论参见 Friedman (1989)；Gaynanova 等人 (2016)。

命题 2.2 和命题 2.3 共同表明，Θ_g 和 Φ_g 会产生相同的投影 QDA 分类器：$h_\Theta(X) = h_\Phi(X)$。因此，在高维情形下，Θ_g 可作为 Φ_g 一种理论上有效且计算稳定的替代方法。实际中，由于 $\widehat{\Sigma}_g^{-1}\widehat{\Sigma}_B$存在奇异性，直接估计 Φ_g并不稳定；而替代量 Θ_g 自然导出式 (4) 中的凸目标函数及其经验对应形式（式 (6)），从而能够可靠地进行稀疏估计和组级变量选择。

根据命题 2.1，令 $\Theta_g := (\Sigma_g + \Sigma_B)^{-1}\Gamma$，可得

$$
\begin{aligned}
\Theta_g &= \arg\min_{\Theta\in\mathbb{R}^{p\times(G-1)}} \frac{1}{2}\|(\Sigma_g+\Sigma_B)^{1/2}\Theta - (\Sigma_g+\Sigma_B)^{-1/2}\Gamma\|_F^2 \\
&= \arg\min_{\Theta\in\mathbb{R}^{p\times(G-1)}} \frac{1}{2}\mathrm{Tr}\big(\Theta^\top(\Sigma_g+\Sigma_B)\Theta - 2\Gamma^\top\Theta\big) \\
&= \arg\min_{\Theta\in\mathbb{R}^{p\times(G-1)}} \frac{1}{2}\mathrm{Tr}(\Theta^\top\Sigma_g\Theta) + \frac{1}{2}\|\Gamma^\top\Theta - I_{(G-1)}\|_F^2, \qquad (4)
\end{aligned}
$$

其中，$I_{(G-1)}$ 表示（$G-1$）$\times$ ($G-1$)单位矩阵。

在实践中，利用从数据中获得的样本估计值 $\widehat{\Sigma}_g$、$\widehat{\Sigma}_B$ 和 $\widehat{\Gamma}$，我们以与其总体对应量类似的方式 Θ_g 定义估计量。为统一符号说明，我们首先介绍组级目标函数中使用的几个矩阵。对于

每个 $g = 1,\ldots,G$，令 $C_{gg} \in \mathbb{R}^{G\times G}$ 表示这样一个矩阵：其第（g,g）个元素等于 1，其余元素均等于 0。定义 $K_g = C_{gg} \otimes I_{(G-1)}$，其中 $\otimes$ 表示克罗内克积，并令 $J_G = \mathbf{1}_G^\top \otimes I_{(G-1)}$。有了这些定义，组级样本判别方向基 Θ_g 可通过以下方式获得

$$\widehat{\Theta}_g = \arg\min_{\Theta_g^* \in \mathbb{R}^{p\times(G-1)}} \left\{ \frac{1}{2}\mathrm{Tr}(\Theta_g^{*\top}\widehat{\Sigma}_g\Theta_g^*) + \frac{1}{2}\|\widehat{\Gamma}^\top\Theta_g^* - I_{(G-1)}\|_F^2 \right\}, \qquad g = 1,\ldots,G.$$

记 $\Theta = [\Theta_1,\ldots,\Theta_G]$，则式 (4) 中的一组 G 优化问题可以等价地写成一个联合目标函数：

$$\Theta = \arg\min_{\Theta_* \in \mathbb{R}^{p\times G(G-1)}} \frac{1}{2}\left\{ \sum_{g=1}^{G} \mathrm{Tr}\left(\Theta_*^\top\Sigma_g\Theta_* K_g\right) + \|\Gamma^\top\Theta_* - J_G\|_F^2 \right\}. \tag{5}$$

相应地，对于样本估计量 Σ_g 和 Γ，记 $\Theta = [\Theta_1,\ldots,\Theta_G]$，我们得到

$$\widehat{\Theta} = \arg\min_{\Theta \in \mathbb{R}^{p\times G(G-1)}} \frac{1}{2}\left\{ \sum_{g=1}^{G} \mathrm{Tr}(\Theta^\top\widehat{\Sigma}_g\Theta K_g) + \|\widehat{\Gamma}^\top\Theta - J_G\|_F^2 \right\}. \tag{6}$$

对 Σ_g 不施加额外约束，尽管当 $p \gg n_g$ 时它可能是奇异的，因为式 (6) 中的目标函数有下界。虽然式 (6) 中的优化问题可用于确定投影基 Θ，但它本身并不能实现变量选择。

为引入稀疏性，我们考虑一种嵌套组结构，以促进组联合和组特异性变量选择。对于每个变量 j，令 $\omega_j \in \mathbb{R}^{G(G-1)}$ 表示 Ω 的第 j 行，并将其视为一个向量，记作 $\omega_j = (\omega_{j1}^\top, \ldots, \omega_{jG}^\top)^\top$，其中每个 $\omega_{jg} \in \mathbb{R}^{G-1}$ 都是与组 g 对应的块。我们的惩罚目标函数为

$$\widehat{\Omega} := \widehat{\Omega}(\lambda,\alpha) = \arg\min_{\Omega \in \mathbb{R}^{p\times G(G-1)}} L, \tag{7}$$

$$L = \frac{1}{2}\left\{ \sum_{g=1}^{G} \mathrm{Tr}(\Omega^\top\widehat{\Sigma}_g\Omega K_g) + \|\widehat{\Gamma}^\top\Omega - J_G\|_F^2 \right\} + \alpha\lambda\sum_{j=1}^{p}\|\omega_j\|_2 + \frac{1-\alpha}{\sqrt{G}}\lambda\sum_{j,g}\|\omega_{jg}\|_2. \tag{8}$$

公式 (8) 是凸的，且有下界。第一项对应于 (6)，通过平衡组内变异性与组间分离度，在样本版本中估计投影基。第二项是稀疏惩罚项，使投影基能够进行变量选择。调参参数 λ 控制整体稀疏程度：λ 值越大，$\Omega(\lambda,\alpha)$ 越稀疏。参数 $\alpha \in (0,1]$ 用于平衡行级稀疏性与组特异性稀疏性：α 越接近 1

就越侧重于行稀疏性，从而鼓励在所有组中同时选择（或排除）变量；而 α 较小时，则更强调块稀疏性，允许针对不同组选择不同变量。在模拟研究和真实数据应用中，我们将 $\alpha = 0.5$ 设定为在这两种稀疏性之间取得平衡。我们在附录 S.3.5 节中给出了 Ω 中共享变量选择与组特异性变量选择的示例。

我们注意到，组特异性投影基的拼接在 MGQDA 中发挥着重要的结构作用。下面的注释说明了为何将各组估计量堆叠起来是恰当的。

注 1. 由于每个组 g 都有自己的类内协方差 Σ_g，我们估计各自独立的投影基 Θ_g，然后通过按列拼接 $\Theta = [\Theta_1, \ldots, \Theta_G]$ 构成联合基（见式 (6)）。这种堆叠保留了组特异性的判别方向，并使惩罚公式-(8) (8)中能够同时进行共享变量选择和组特异性变量选择。取平均 $\{\Theta_g\}$ 会抹去异质性、模糊不同的方向，并与惩罚项的组块结构相冲突。由于投影 *QDA* 规则仅取决于所收集方向张成的空间，堆叠既保持了分类不变性，也符合我们的理论保证（定理 *2–3*）。

2.4 优化算法

我们采用块坐标下降算法（Tseng，2001）求解式（7）。由于式（8）中的目标函数是凸的，任何极小化解都必须满足 Karush-Kuhn-Tucker（KKT）条件（Boyd 和 Vandenberghe，2004）。对 Ω 中每个 $(G-1) \times 1$ 向量 $\widehat{\omega}_{jg}$，分别关于求式（8）中目标函数的偏导数，其中 $j = 1, \ldots, p$ 以及 $g = 1, \ldots, G$，得到

$$\widehat{\Sigma}_{g,jj}\widehat{\omega}_{jg} + (\widehat{\Gamma}\widehat{\Gamma}^\top)_{jj}\widehat{\omega}_{jg} + \frac{\alpha\lambda\widehat{\omega}_{jg}}{\|\widehat{\omega}_j\|_2} + \frac{(1-\alpha)\lambda\widehat{\omega}_{jg}/\sqrt{G}}{\|\widehat{\omega}_{jg}\|_2} = -\mathbf{v}_{jg}, \tag{9}$$

其中，A_{ij} 表示矩阵 A 中位于第 i 行、第 j 列的元素，而$\mathbf{v}_{jg}$ 则通过对式（8）中的 $\widehat{\omega}_{jg}$ 项求偏导数得到：

$$\mathbf{v}_{jg} = \sum_{i \neq j} \widehat{\Sigma}_{g,ji}\widehat{\omega}_{ig} + \sum_{i \neq j} (\widehat{\Gamma}\widehat{\Gamma}^\top)_{ji}\widehat{\omega}_{ig} - \widehat{\Gamma}_j. \tag{10}$$

每当 $\widehat{\omega}_{jg} \neq 0$ 时，式 (9) 左侧的每一项都与 $\widehat{\omega}_{jg}$ 成正比。因此，KKT 条件意味着 $-\mathbf{v}_{jg}$ 与 $\widehat{\omega}_{jg}$ 共线。于是存在一个

^标量 $\vartheta_{jg} > 0$，它取决于 $\|\widehat{\omega}_{jg}\|_2$ 和 $\|\widehat{\omega}_j\|_2$，使得 $\mathbf{v}_{jg} = -\vartheta_{jg}\widehat{\omega}_{jg}$，$\vartheta_{jg} = \Sigma_{g,\ jj} + (\widehat{\Gamma}\widehat{\Gamma}^\top)_{jj} + \frac{\alpha\lambda}{\|\widehat{\omega}_j\|_2} + \frac{1-\alpha}{\sqrt{G}}\frac{\lambda}{\|\widehat{\omega}_{jg}\|_2}$.。令 $x = \|\widehat{\omega}_{jg}\|_2$、$a = \widehat{\Sigma}_{g,\ jj} + (\widehat{\Gamma}\widehat{\Gamma}^\top)_{jj}$、$b = \left(\sum_{t\neq g}\|\widehat{\omega}_{jt}\|_2^2\right)^{1/2}$ 和 $c = \|\mathbf{v}_{jg}\|_2$。对式 (9) 两边取范数，在 $x > 0$ 时该操作有效，由此得到标量方程 $ax + \alpha\lambda\frac{x}{\sqrt{b^2+x^2}} + \frac{1-\alpha}{\sqrt{G}}\lambda = c.$。由于 $\frac{d}{dx}\left[x/\sqrt{b^2+x^2}\right] = b^2(b^2+x^2)^{-3/2} \geq 0$，左侧关于 $x \geq 0$ 严格递增。因此，当 $c > \frac{1-\alpha}{\sqrt{G}}\lambda$ 时，存在唯一解 $x > 0$；否则，块更新为 $x = 0$（次梯度情形）。可通过牛顿法或二分法高效求根；一个方便的搜索区间是 $x \in \left[0,\ \left(c - \frac{1-\alpha}{\sqrt{G}}\lambda\right)_+/a\right]$。详细算法见附录 S.3.5 节。

3 理论结果

在本节中，我们证明，在正态性和不可表示性条件下，并满足协方差偏差的适当界限及最小信号强度要求，所提方法能够实现变量选择一致性。对于某个与 (n, p) 无关的常数 $C > 0$，若 $a_n \leq Cb_n$，则记为 $a_n \lesssim b_n$；对于某个这样的 C，若 $a_n \geq Cb_n$，则记为 $a_n \gtrsim b_n$。若 $a_n \lesssim b_n$ 且 $a_n \gtrsim b_n$，则记为 $a_n \asymp b_n$。

^ ^ ~ ~对于 Γ 的每一列 Γ_r，令 Σ_r 表示其协方差矩阵，并令 $\Sigma_{r,\ s}$ 表示列 Γ_r 与 Γ_s 之间的^ ^协方差。预言投影基 Θ 可分解为 $\Theta = (\Theta_1, ..., \Theta_G)$，如第 2.3 节所示；对于每个 $g \in \{1, \ldots, G\}$，我们有 $\Theta_g = (\Sigma_g + \Sigma_B)^{-1}\Gamma_{\cdot g}$。$\Theta$ 的支撑集定义为 $S = \{j \in [p] : \|\Theta_j\|_2 \neq 0\}$，其中 Θ_j 表示 Θ 的第 j 行。对于每个组 $g = 1, ..., G$，令 $S_g = \{j \in [p] : \|\Theta_{g,\ j}\|_2 \neq 0\}$，并用 $S_g^c = [p] \setminus S_g$ 表示其补集，其中 $\Theta_{g,\ j}$ 表示组 g 中对应于第 j 行的子行向量。因此，$S = S_1 \cup S_2 \cup \cdots \cup S_G$。$S$ 和 S_g 的基数分别记为 $s = |S|$ 和 $s_g = |S_g|$。在实践中，支撑集 S_g 是 $\{1, \ldots, p\}$ 的子集，表示对组 g 的分类结果有贡献的变量；而支撑集 S 也是 $\{1, \ldots, p\}$ 的子集，表示对各组总体分类结果有贡献的变量集合。类似地，^ ^ ^ ^对每一行 $j = 1, ..., p$，我们定义 $S = \{j \mid \|\Omega_j\|_2 \neq 0\}$ 和 $S_g = \{j \mid \|\Omega_{g,\ j}\|_2 \neq 0\}$。为建立变量选择一致性，我们作出以下假设：

(A1) 正态性：我们假设 $\{X \mid Y = g\} \sim \mathcal{N}(\mu_g,\ \Sigma_g)$，$\Pr(Y = g) = \pi_g$ 对于 $g = 1, ..., G$，其中 $\{\max_{g\in\{1,...,G\}}(\pi_g)\}/\{\min_{g\in\{1,...,G\}}(\pi_g)\} = \mathcal{O}(1)$。

(A2) **不可表示性条件：**对于 $g \in \{1, ..., G\}$，存在一个常数 $\psi \in (0, 1]$ 使得 $\left\|\left|\Sigma_{g,S_g^C S_g}\Sigma_{g,S_g S_g}^{-1}\right|\right\|_{\infty,2} \leq 1 - \psi.$

(A3)**协方差偏差界：** 对于 $g \in \{1,...,G\}$和 $r \in \{1,...,G-1\}$， $\left|\left|\left|\widetilde{\Sigma}_{r,S_gS_g}\Sigma^{-1}_{g,S_gS_g}\right|\right|\right|_2 \leq \mathcal{O}(s_g)$; $\left|\left|\left|\Gamma^\top_{S_g}\Sigma^{-1}_{g,S_gS_g}\widetilde{\Sigma}_{r,S_gS_g}\Sigma^{-1}_{g,S_gS_g}\Gamma_{S_g}\right|\right|\right|_2 \leq \mathcal{O}(s_g)$.

(A4) 最小信号强度： 对于 $g \in \{1,...,G\}$，存在常数 $M_g > 0$，使得

$$\text{that } \Theta_{g,\min} = \min_{j\in S_g} \|e_j^\top \Theta_{g,S_g}\|_2 \geq \lambda \left|\left|\left|(\Sigma_{g,S_gS_g} + \Gamma_{S_g}\Gamma^\top_{S_g})\right|\right|\right|_2 \times \left(1+M_g\left[\left|\left|\left|\Gamma^\top_{S_g}\Sigma_{g,S_gS_g}\Gamma_{S_g}\right|\right|\right|_2 \vee 1\right]\times \left[1+\sqrt{(\Sigma_{g,S_gS_g})_{jj} \vee \max_j(\Sigma^{-1}_{g,S_gS_g}\widetilde{\Sigma}_{r,S_gS_g}\Sigma^{-1}_{g,S_gS_g})_{jj}\frac{(G-1)\log\{s_g\log(n_g)\}}{n}}\right]\right).$$

假设 (A1) 是判别分析中的标准假设（Gaynanova 和 Kolar，2015；Mai 等人，2012）。我们不要求对所有 g都有 $\pi_g = 1/G$，但假设各组的组先验概率处于同一数量级。假设 (A2) 是不可表示性的一种形式，在套索回归类方法的变量选择过程中常见于理论分析（Zhao 和 Yu，2006；Kolar 和 Liu，2014；Obozinski 等人，2011），并且是确保正确恢复支撑集所必需的。假设 (A3) 通过要求总体层面存在温和的分离，控制协方差相关量及其替代量之间的差异；这种分离使组级二次型在投影后仍可区分。这确保了判别方向的可辨识性以及投影分类器的稳定性，是多类高维判别分析中常用的一类条件（例如，Fukunaga（1990）、Mai 等人（2012）、Mai 等人（2019年））。假设 (A4) 要求每个相关变量都携带足够强的信号，以压过噪声，这与 Gaynanova 和 Kolar（2015）中使用的 beta-min 条件相类似。

$\widehat{}$为了建立支撑集 S_g 与 S_g之间的变量选择一致性，我们采用一种常见策略，引入一个连接式（5）中定义的 Θ、$\widehat{}$与式（7）中定义的 Ω 的中间"桥接"估计量。具体而言，我们考虑一个由总体协方差矩阵构造、且包含相同惩罚结构的总体投影基：

$$\Omega = \underset{\Omega_* \in \mathbb{R}^{p\times G(G-1)}}{\arg\min} \left\{ \frac{1}{2}\left(\sum_{g=1}^{G} \text{Tr}(\Omega_*^\top \Sigma_g \Omega_* K_g) + \|\Gamma^\top \Omega_* - J_G\|_F^2 \right) + \lambda\left(\alpha \sum_{j=1}^{p} \|\omega_{j*}\|_2 + \frac{(1-\alpha)}{\sqrt{G}} \sum_{j=1}^{p}\sum_{g=1}^{G} \|\omega_{jg*}\|_2 \right) \right\}, \tag{11}$$

其中， ω_{j*} 表示与 Ω_* 的第 j 行对应的长度为 $G(G-1)$ 的行向量，而 ω_{jg*} 是与组 g 相关联的其 $(G-1)$ 维子向量，其定义方式与总体矩阵 Ω 中的 ω_j 和 ω_{jg} 类似。令 u_{S_g} 表示与 $\Omega_{g,\ S_g}$相关的惩罚项的次梯度项，定义为 $u_{S_g} = \sum_{j\in S_g}\left\{\alpha \frac{\omega_{jg}}{\|\omega_j\|_2} + (1-\alpha)\frac{\omega_{jg}}{\|\omega_{jg}\|_2}\right\}$， $g =$ —— ——

1，……，G. 下一个结果刻画了 Θ 与总体惩罚解 Ω 之间的关系。

定理 *1.* 对于每个组 $g = 1, ..., G$，假设 $\left\|\Sigma_{g,S_g^c S_g}\Sigma_{g}^{-1},_{S_g S_g} u_{S_g}\right\|_{\infty,2} < 1$，且调参参数 $\lambda < \Theta_{g,\min}\left(\left\|(\Sigma_{g,S_g S_g} + \Gamma_{S_g}\Gamma_{S_g}^{\top})^{-1}\right\|_{\infty}\left(1 + \left\|\Gamma_{S_g}^{\top}\Sigma_{g,S_g S_g}^{-1}\Gamma_{S_g}\right\|_{2}\right)\right)^{-1}$。那么，$\Omega_g$ 的解具有 $\Omega_g = (\Omega_{g}^{\top},_{S_g}, 0_{p-s_g}^{\top})^{\top}$ 的形式，其中 $\Omega_{g,S_g} = \Theta_{g,S_g}(I + \Gamma_{S_g}^{\top}\Sigma_{g}^{-1},_{S_g S_g}\Gamma_{S_g})^{-1} - \lambda(\Sigma_{g,S_g S_g} + \Gamma_{S_g}\Gamma_{S_g}^{\top})^{-1}u_{S_g}$。此外，对于所有 $j \in S_g$，$\|e_j^{\top}\Omega_{g,S_g}\|_2 \neq 0$。

该定理表明，在满足不可表示性条件且适当选择惩罚参数 λ的情况下，对于每个 $g = 1, \ldots, G$，总体惩罚解 Ω_g 与未惩罚解 Θ_g 具有相同的支撑集。该定理还给出了 Θ_g 与 Ω_g之间的明确关系，我们将在下一个定理中用它推导样本估计量 Ω 的变量选择一致性^。

定理 2. 在假设（*A1*）–（*A4*）下，若样本量满足 $n \gtrsim \mathcal{O}(s\log\, p)$，且调参参数的选择满足 $\lambda \gtrsim (\log p)/n$，则式 (7) 中定义的 *MGQDA*估计量^具有变量选择一致性；也就是说，对于所有 $g = 1, \ldots, G$， $S_g = S_g$均成立，且该结论以至少 $1 - \mathcal{O}(\log^{-1} n)$的概率成立。

该定理意味着，随着样本量增加，该方法以趋近于 1 的概率正确识别每个组中的真实非零变量，从而在组层面确立变量选择一致性-^。此外，对于堆叠估计量^Ω的活跃块集 S，我们得到如下结果。

推论1。 在与定理 *2*相同的假设下，如果对于每个组 $g = 1$、……、G、^ ^ ^ ^的堆叠估计量 Ω $= [\Omega_1$、……、$\Omega_G]$，我们以至少 $1 - \mathcal{O}(\log^{-1}(n))$ 的概率有 $S = S$。

^也就是说，Ω 的活跃块与 Ω 的活跃块一致，从而在聚合层面建立了变量选择一致性。此外，在满足关于样本量和数据结构的适当假设时，我们的方法保证基于样本得到的分类结果与底层总体分布得到的分类结果一致。具体而言，对于每个组 g，定义 κ_g: $= \left\|\Sigma_{g,S_g^C S_g}\Sigma_{g}^{-1},_{S_g S_g}\right\|_{\infty},\Lambda_g := \Theta_{S_g}^{\top}\Sigma_{g,S_g S_g}\Theta_{S_g} \in \mathbb{R}^{(G-1)\times(G-1)}$ 并令 $\lambda_{\min}(\Lambda_g)$ 表示 Λ_g 的最小特征值。

定理 3。 假设（*A1*）成立。满足以下条件：（***B1***） $n, p \to \infty$, $G = \mathcal{O}(1)$, $s^2 \log(ps^2)/n \to 0$, 以及 $\mathcal{O}\left(\sqrt{s^2 \log(ps^2)/n}\right) < \mathcal{O}(\lambda) < \mathcal{O}\left(\min_g \Theta_{g,\min}\right)$; （***B2***） $\inf_g \lambda_{\min}(\Lambda_g) > 0$。若对于每个 $g = 1, \ldots, G$, 均有 $\lambda \to 0$ 且 $\kappa_g < 1$, 则

$$P\left(\hat{h}_{\widehat{\Omega}}(X) = h_{\Theta}(X)\right) \to 1,$$

其中 $h_{\Theta}(X_{new}) = \arg\min_{1 \le g \le G}\{(X_{new} - \mu_g)^\top \Theta \left(\Theta^\top \Sigma_g \Theta\right)^{-1} \Theta^\top (X_{new} - \mu_g) + \log\left|\Theta^\top \Sigma_g \Theta\right| - 2\log \pi_g\}$ 以及 $\hat{h}_{\widehat{\Omega}}(X_{new}) = \arg\min_{1 \le g \le G}\{(X_{new} - \bar{X}_g)^\top \widehat{\Omega}(\widehat{\Omega}^\top \widehat{\Sigma}_g \widehat{\Omega})^{-1} \widehat{\Omega}^\top (X_{new} - \bar{X}_g) + \log\left|\widehat{\Omega}^\top \widehat{\Sigma}_g \widehat{\Omega}\right| - 2\log(n_g/n)\}$.

注 2。 在相同假设下，该估计量在活跃集 $S = \cup_g S_g$ 上也具有分块一致性，即 $\left\|\left|(\widehat{\Omega} - \Theta)_S\right\|\right|_{\infty, 2} \to 0$，这一结论可由定理 *3* 证明中的 ***KKT*** 条件和高概率界推出。由于分类器仅通过投影基依赖于基，μ_g 和 Σ_g（对于 $g = 1, \ldots, G$），这保证了二次项和对数行列式项一致收敛，从而得到上述分类一致性。

该定理表明，当样本量满足 $n \gtrsim s^2 \log(ps^2)$ 时，所提方法同时实现变量选择一致性和分类一致性；也就是说，基于样本估计值的分类器会收敛于总体层面的分类器。除了高斯且满秩的情形外，我们还考虑了两种扩展。首先，我们分析样本投影基 Ω 不满秩的情形，并明确给出相应的分类器。令 $A_g := \widehat{\Omega}^\top \widehat{\Sigma}_g \widehat{\Omega}$。如果 A_g 可逆，则所提方法按照公式 (3) 使用 A_g^{-1} 和 $\log|A_g|$。如果 A_g 奇异，则改用 Moore–Penrose 伪逆 A_g^{+} 以及伪行列式 $\mathrm{pdet}(A_g)$，其定义为正特征值的乘积。其次，我们研究理论结果扩展到次高斯总体所需的条件，并表明在适当假设下，所提方法仍保持相同的一致性保证。定理 1–3 的详细证明，以及对这两种扩展的讨论，见附录 S.1–S.3 节。

4 模拟研究

4.1 模拟设置

我们从分类准确率和变量选择两个方面评估 MGQDA 的性能。预测准确率通过误分类率进行衡量，变量选择则使用真正率（TPR）和假正率（FPR）进行评估。

我们将 MGQDA 与以下方法进行比较：基于收缩均值的对角二次判别分析（SMDQDA，Tong 等人，2012）、高维正则化判别分析（HDRDA，Ramey 等人，2016）、多组稀疏判别分析（MGSDA，Gaynanova 等人，2016）、投影判别分析（DAP，Gaynanova 和 Wang，2019）、稀疏线性判别分析（SLDA，Clemmensen 等人，2011）以及惩罚线性判别分析（pLDA，Witten 和 Tibshirani，2011）。由于 DAP 是一种两组方法，我们通过成对投票将其扩展到多组问题，并将其记为“GDAP”。最初曾考虑基于协方差矩阵广义逆的经典二次判别分析，但由于数值不稳定性和错误率偏高，最终将其排除在比较之外。

在所有模拟中，我们将每个 g的组特定样本量设为 $n_g = 100$ ，考虑 $p \in \{200, 500\}$ 个预测变量，并考察 $G \in \{3, 5\}$ 个组。数据来自多元正态分布（$X \mid Y = g$） $\sim \mathcal{N}$（μ_g， Σ_g）。为涵盖一系列协方差结构，我们考虑四类结构：(i) 分块等相关结构，块大小为b ，参数为 $\rho \in [0, 1]$，其中 $\Sigma_g = \text{diag}\{\rho I_b + (1-\rho)\iota_b\iota_b^\top$， $I_{p-b}\}$ ，且 ι_b 为 b维全1向量；(ii) 分块自相关结构，块大小为 b ，参数为 $\rho \in [0, 1]$，其中$\Sigma_g = \text{diag}\{\widetilde{\Sigma}_b, I_{p-b}\}$ 且 $(\widetilde{\Sigma}_b)_{ij} = \rho^{|i-j|}$；(iii) 尖峰协方差结构，其归一化向量为$q_1$、$q_2 \in \mathbb{R}^p$，标量为 a_1、 $a_2 \in \mathbb{R}$，其中 $\Sigma_g = a_1 q_1 q_1^\top + a_2 q_2 q_2^\top + I_p$；除非另有说明，我们设 $q_1 \propto$(1，...， b，0，...，0)，并设 $q_2 \propto$(b，...，1，0，...，0)；(iv) 分块模型，其中令 $\Lambda =$diag(λ_1，...， λ_b)，其中 $\lambda_i \sim$Unif$[1, 2]$ ，且 U为一个元素独立同分布且服从标准正态分布的 $b \times b$矩阵，并设 $\Sigma_g = \text{diag}\{U^\top \Lambda U, I_{p-b}\}$。我们考虑八个数据生成模型。模型1–5使用 $G = 3$ 个组，模型6–8使用 $G = 5$ 个组。对于每个模型，我们考察$p = 200$ 和 $p = 500$。对于模型1–4，我们设 $b = 30$ ；对于模型5–8，我们设 $b = 50$ 。均值与协方差结构如下：

模型 1： $\mu_1 = \mathbf{0}_p$, Σ_1 为分块等相关结构， $\rho = 0.8$； $\mu_2 =$($\mathbf{1}_{10}, -\mathbf{1}_{10}, \mathbf{0}_{p-20}$)， Σ_2 为分块自相关结构， $\rho = 0.8$； $\mu_3 =$ ($-\mathbf{1}_{10}$,$\mathbf{1}_{10}, \mathbf{0}_{p-20}$)， Σ_3 具有尖峰结构，参数为（a_1, a_2） =(100, 10)。

模型 2： $\mu_1 = \mathbf{0}_p$； $\mu_2 =$ ($\mathbf{0.5}_{10}, -\mathbf{0.5}_{10}, \mathbf{0}_{p-20}$)； $\mu_3 =$ ($-\mathbf{0.5}_{10}$,$\mathbf{0.5}_{10}, \mathbf{0}_{p-20}$)；协方差矩阵（$\Sigma_1, \Sigma_2, \Sigma_3$）与模型 1 中的相同。

模型 3： $\mu_1 = \mathbf{0}_p$, Σ_1 为分块等相关结构， $\rho = 0.3$； $\mu_2 =$ ($\mathbf{1}_{10}, -\mathbf{1}_{10}, \mathbf{0}_{p-20}$)， Σ_2 为分块自相关结构， $\rho = 0.7$； $\mu_3 =$ ($-\mathbf{1}_{10}$,$\mathbf{1}_{10}, \mathbf{0}_{p-20}$)， Σ_3 来自分块模型，其中U, Λ 按上述方式生成。

模型 4： $\mu_1 =$($\mathbf{1}_{10}$， $\mathbf{0}_{p-10}$)， Σ_1 为分块等相关，且 $\rho = 0.3$； $\mu_2 =$ （$\mathbf{0}_{10}$，$\mathbf{1}_{10}$，$\mathbf{0}_{p-20}$）， Σ_2 为

分块自相关，且 $\rho = 0.7$； $\mu_3 =$ （$\mathbf{0}_{20}$，$\mathbf{1}_{10}$，$\mathbf{0}_{p-30}$），Σ_3 具有尖峰结构，参数为（a_1，a_2）=（30，5），这里设 $q_1 \propto$（1，…，$\sqrt{b}$，0，…，0）和 $q_2 \propto$（$\sqrt{b}$，…，1，0，…，0）。

模型5： 令 $\rho_0 = 0.8$。则 $\mu_1 = (\mathbf{1}_{10}, \mathbf{0}_{p-10})$，$\Sigma_1$ 为分块自相关，参数为 $\rho = \rho_0$； $\mu_2 = (\mathbf{0}_{10}, \mathbf{1}_{10}, \mathbf{0}_{p-20})$，$\Sigma_2$ 为分块自相关，参数为 $\rho = 0.7\rho_0$； $\mu_3 = (\mathbf{0}_{20}, \mathbf{1}_{10}, \mathbf{0}_{p-30})$，$\Sigma_3$为分块自相关，参数为 $\rho = 0.3\rho_0$。

模型6： $\mu_1 = (110, \mathbf{0}_{p-10})$，$\Sigma_1$ 为分块等相关，参数为 $\rho = 0.5$； $\mu_2 = (010, 110, \mathbf{0}_{p-20})$，$\Sigma_2$ 为分块自相关，参数为 $\rho = 0.5$； $\mu_3 = (020, 110, \mathbf{0}_{p-30})$，$\Sigma_3$ 为尖峰结构，参数为 $(a_1, a_2) = (100, 10)$； $\mu_4 = (030, 110, \mathbf{0}_{p-40})$，$\Sigma_4$ 为尖峰结构，参数为 $(a_1, a_2) = (10, 100)$； $\mu_5 = (040, 110, \mathbf{0}_{p-50})$，$\Sigma_5$ 来自分块模型，参数为 U，Λ 按上述方式生成。

模型7： 协方差矩阵与模型6相同。均值向量按比例缩小： $\mu_g^{(7)} = 0.5\,\mu_g^{(6)}$适用于 $g = 1$、……、5，其中 $\mu_g^{(6)}$表示模型6中的均值。

模型8： 协方差矩阵再次与模型6相同。均值向量为$\mu_1 = \mathbf{0}_p$； $\mu_2 = (\mathbf{1}_{20}, \mathbf{0}_{p-20})$； $\mu_3 = (\mathbf{0.2}_{10}, -\mathbf{0.2}_{10}, \mathbf{0}_{p-20})$； $\mu_4 = (\mathbf{0}_{20}, \mathbf{5}_2, \mathbf{0}_{p-22})$； $\mu_5 = (\mathbf{0}_{25}, -\mathbf{10}, \mathbf{10}, \mathbf{0}_{p-27})$。

对于每种方法，使用独立测试集上的经验分类错误率来衡量预测准确率。变量选择指标（真正率/假正率）和组特异性活动集的计算细节见第4.2节和附录S.5节。

4.2 结果

总体而言，在几乎所有模拟设置中，MGQDA 的表现都很出色，在所比较的方法中通常能取得最低或接近最低的误分类率（见图1）。对比模型1和模型2（以及模型6和模型7）可以发现，即使均值差异相对较弱，MGQDA、HDRDA、GDAP 和 SMDQDA 仍具竞争力，这表明这些方法能够有效利用协方差结构的差异。相比之下，当各类别的组内协方差矩阵变得更加同质时（例如模型5），完全二次判别规则的优势自然会减弱，而所有方法的相对表现也会变得更加相似。

图 2 和图 3 汇总了稀疏方法（MGQDA、MGSDA、pLDA 和 SLDA）的整体变量选择性能，报告了 100 次重复试验中的平均真正率和假正率。MGQDA 在几乎所有模拟设置中都实现了灵敏度与

特异性之间最理想的平衡。随着维度从 $p = 200$ 增至 $p = 500$，其真正率始终保持较高且稳定，而假正率则始终较低，表明该方法能够准确识别活跃变量，同时避免过多的误选。

相比之下，MGSDA 的 TPR 处于中等水平，但变异性明显更大，FPR 也更高，反映出其支撑集估计的稳定性较差。pLDA 往往能够达到相对较高的 TPR，但 FPR 显著膨胀，这表明尽管检测出了许多真正的信号，同时也纳入了许多噪声变量。SLDA 倾向于生成过于稀疏的模型，导致 TPR 极低，判别效用有限。综合这些比较结果可以看出，在异质协方差结构下，MGQDA 能够提供最稳定、最可靠的支撑集恢复。

为进一步考察组层面的变量选择，附录 S.5 提供了详细的特定于组的 TPR 和 FPR 结果，并明确构造了真实活跃集和估计活跃集。以设置 3 为例，附录 S.5 展示了如何$\widehat{\ }$为每个组构建支撑集 S_g 和 S_g，并表明 MGQDA 在异质组之间保持了稳健的性能，这与图 2–3 中报告的汇总结果一致。

结合误分类结果，这些发现表明，在所有模型、维度和协方差结构下，MGQDA 始终在预测准确性与变量选择可靠性之间实现了最优的组合。这些实证证据支持第 3 节中的理论保证。

5 真实数据应用

我们使用 DepMap Public 23Q2 版本中的基因表达数据评估所提出方法的性能（DepMap、Broad，2023；Arafeh 等，2025）。该数据集提供了癌细胞系的大规模基因组和功能特征，广泛用于研究癌症中的遗传依赖性和脆弱性（Killian 和 Gatto，2021；Bazaga 等，2020；Chiu 等，2021）。该数据集包含 1,864 个癌细胞系基于 RNA测序的表达测量值，以 $\log_2(\text{TPM} + 1)$ 变换后的形式报告。总体而言，共有 19,193 个蛋白质编码基因的表达水平可用，为评估所提出的分类器提供了全面且标准化的资源。

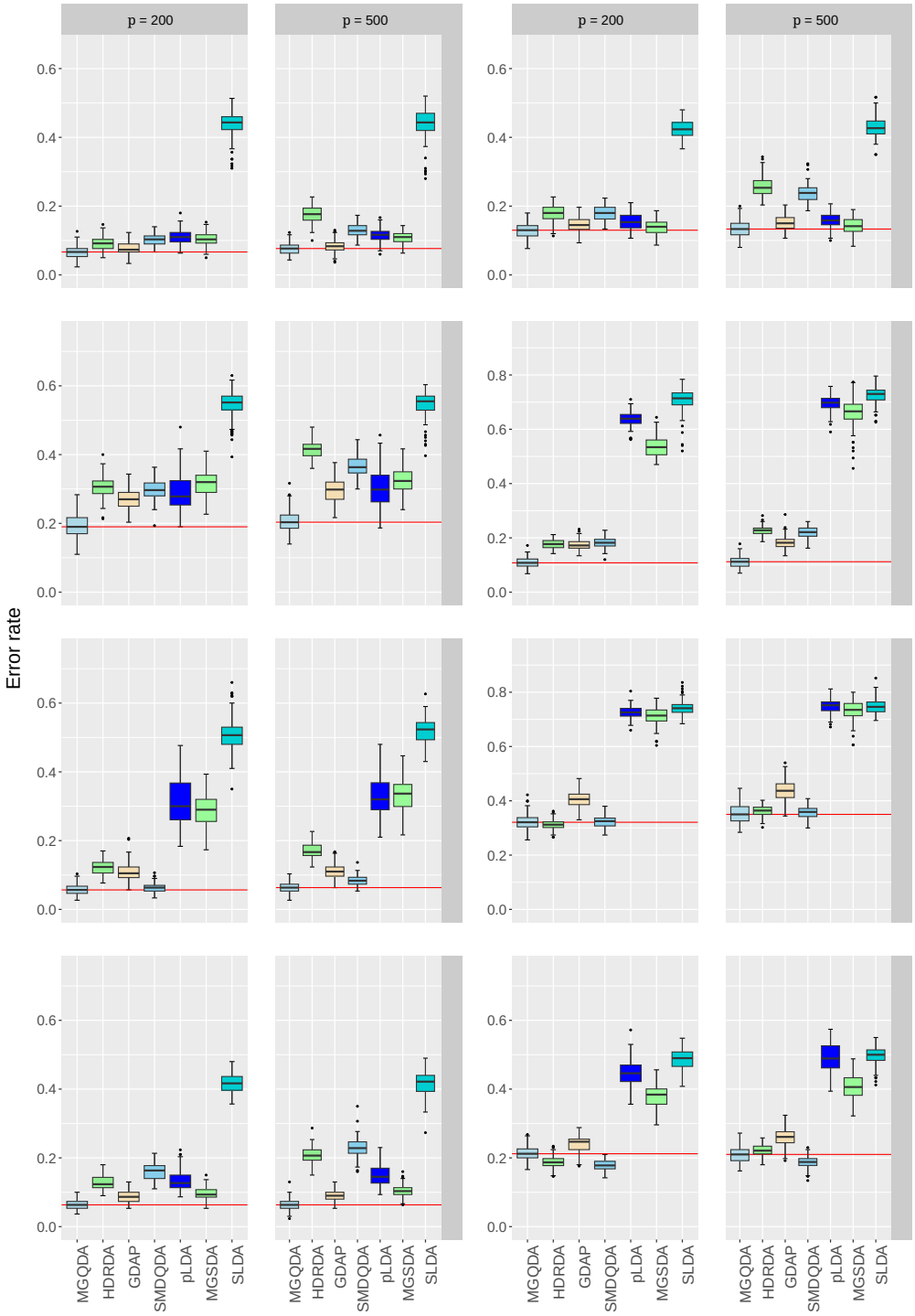

图1： 基于100次重复试验，MGQDA与竞争方法在所有模拟设置下的误分类错误率。每个面板中的水平线表示MGQDA的中位数错误。

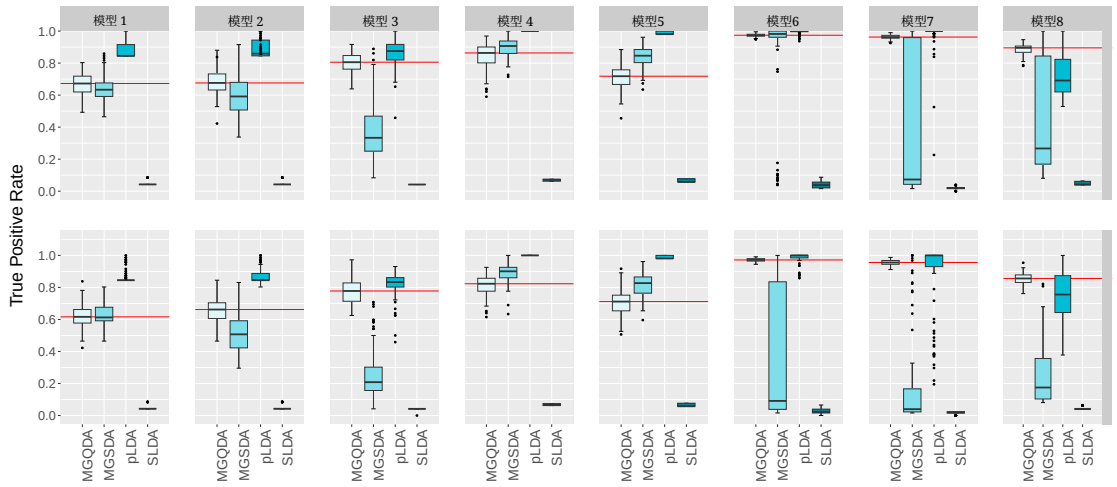

图2： MGQDA、MGSDA、pLDA 和 SLDA 在所有模拟设置下的真正率，基于100次重复试验。每个面板中的红色水平线表示 MGQDA 的真阳性率中位数。

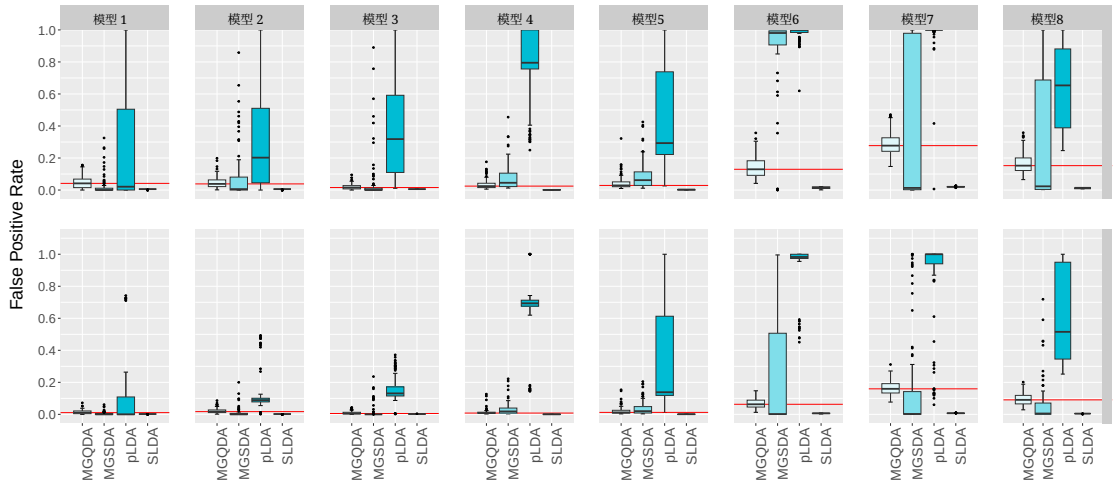

图3： MGQDA、MGSDA、pLDA 和 SLDA 在所有模拟设置下的假正率，基于100次重复试验。每个面板中的红色水平线表示 MGQDA 的假阳性率中位数。

由于数据维度较高，我们首先进行了成对 t 检验，以筛选重要基因，并保留 $p = 800$ 个基因用于后续分析。Gaynanova 等人 (2016)、Witten 和 Tibshirani（2011）以及 Gaynanova 和 Wang（2019）也采用了类似的筛选策略。随后，我们将 MGQDA 与其他多组判别分析方法进行了基准比较，包括 SMDQDA、HDRDA、MGSDA 和 pLDA。由于 SLDA 在我们的模拟研究中表现明显较差，因此将其排除在外；此外，我们未在此处纳入 DAP，因为它本质上是一种两类方法（为完整起见，我们仅在模拟中使用了成对投票扩展“GDAP”）。

在主要分析中，我们聚焦于原发病灶样本数量最多的五种疾病：“非小细胞肺癌”“非霍奇金淋巴瘤”“头颈部鳞状细胞癌”“弥漫性胶质瘤”和“结直肠腺癌”。每种疾病包含 40 至 77 个样本。为便于直接比较，我们在所有方法中固定了 $p = 800$ 。为评估稳健性，我们还使用不同的疾病子集和不同的 p 值进行了另外三项实验；详细结果见附录 S.4。

图 4 总结了 100 次重复试验中的表现。在每次重复试验中，90% 的样本被随机分配到训练集，其余 10% 分配到测试集。分类器利用训练集中的遗传数据和疾病标签进行训练，并在留出的测试集上进行评估。所提出的 MGQDA 方法取得了低于竞争方法的分类错误率（图 4，左图）。就误分类而言，HDRDA 的表现相当，但不进行变量选择。

关于变量选择结果，MGQDA 一贯选择比 MGSDA 稀疏得多的模型（图 4，中间），且相当一部分特征仅在特定组中被选中，凸显了我们的方法识别结构差异的能力。HDRDA、SMDQDA 和 pLDA 使用全部 p 变量，因此未纳入变量选择图。在 MGQDA 最常选中的基因中，有几个

得到了现有文献的有力支持。在“非小细胞肺癌”组中，DDR2、PLEK2 和 CDH13 均是入选率最高的基因之一，这与 Sato 等（1998）、Shen 等（2019年）和 Hammerman 等（2011）分别报告的这些基因的作用一致。在“弥漫性胶质瘤”组中，RAB34 经常被选中，这与其在胶质瘤生物学中的已知相关性相符（Wang 等，2015）。

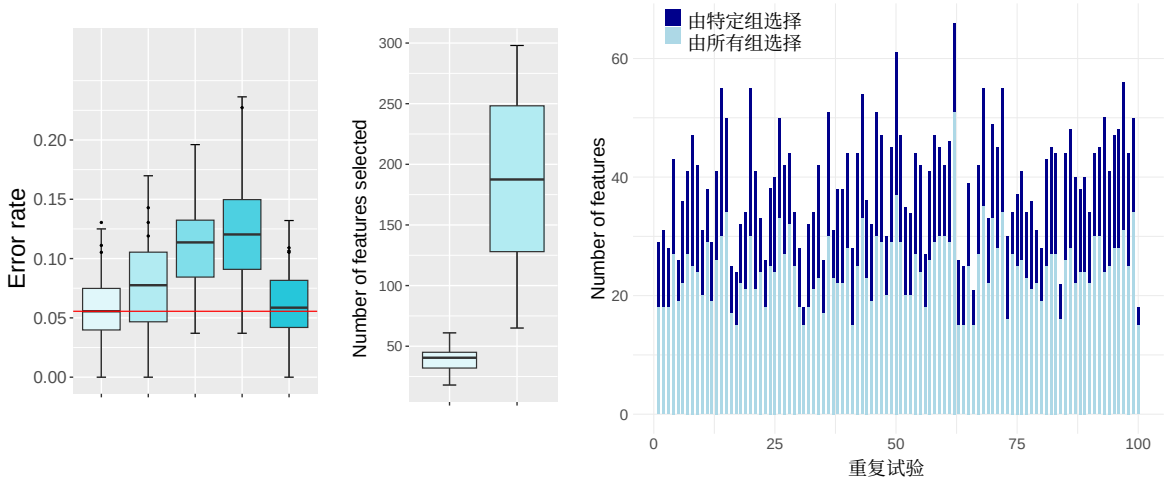

图 4： MGQDA 与竞争方法在 DepMap 基因表达数据集上的分类性能和变量选择。左图：100 次重复试验中的分类错误率（红色水平线：MGQDA 中位数错误）。中图：稀疏方法在各次重复试验中选定的变量数量。右图：按实验划分的 MGQDA 选定变量数量； x 轴表示实验， y 轴表示投影基中选定的变量数量。

因此，MGQDA 在选择能够捕捉共享结构和组特异性结构的稀疏变量集的同时，实现了更低的分类错误率。这带来了具有可解释性的分类规则，并且现有癌症基因组学研究支持了排名靠前的选定基因的生物学相关性。

6 讨论

我们提出了一种新的多组判别分析方法，既能适应不同组之间各异的协方差结构，又能保留有效的变量选择能力和实用的计算成本。大量模拟研究（第 4 节）和真实数据分析（第 5 节）表明，在

具有异质协方差矩阵的情形下，MGQDA 能显著提高分类准确率。除了预测性能有所提升，该方法在计算上仍然高效；当 G 和 n 固定时，其复杂度随 p 线性增长，因此比依赖直接进行高维协方差估计的流程更实用。我们框架的一项主要优势是能够进行组特异性变量选择，DepMap 分析便展示了这一点：MGQDA 可以识别共有标志物和疾病特异性标志物，且结果易于解释。一项自然的扩展方向是借鉴正则化判别分析的思路，纳入收缩策略，以便在部分组的样本量极小时提高稳健性。我们将这些拓展以及 MGQDA 在大规模研究中的更广泛应用留待未来研究。

补充材料

附录：符号说明和技术预备知识；所有命题和定理（定理 1–3）的详细证明及辅助引理；实验设置的补充说明；以及数据分析的补充结果。

R 代码：本文提出的 MGQDA 的实现 R 代码 .

数据可用性声明：本研究分析的 DepMap Public 23Q2 数据（figshare DOI: 10.6084/m9.figshare.22765112.v2）可通过 DepMap 门户网站（ https://depmap.org/portal）公开获取。该数据集仅包含细胞系级别的测量数据，不包含个人身份信息。

参考文献

Arafeh, R., T. Shibue, J. M. Dempster, W. C. Hahn 和 F. Vazquez（2025）。癌症依赖图谱的现状与未来。《自然综述：癌症》*25*(1), 59–73。

Bazaga, A., D. Leggate 和 H. Weisser（2020）。用于预测肿瘤学中新型治疗靶点的基因-癌症关联全基因组调查。《科学报告》*10*(1), 10787。

Blasco, H., J. B-laszczy´nski, J.-C. Billaut, L. Nadal-Desbarats, P.-F. Pradat, D. Devos, C. Moreau, C. R. Andres, P. Emond, P. Corcia 等（2015）。靶向代谢组学的比较分析：基于优势关系的粗糙集方法与正交偏最小二乘判别分析的比较。*Journal of Biomedical Informatics 53*, 291–299。

Boyd, S. P. 和 L. Vandenberghe（2004）。凸优化。剑桥大学出版社。

Cai, T. T. 和 L. Zhang（2021）。一种用于高维稀疏二次判别分析的凸优化方法。*The Annals of Statistics 49*（3）, 1537–1568。

Chen, R.-C., C. Dewi, S.-W. Huang 和 R. E. Caraka（2020）。基于机器学习方法选择用于数据分类的关键特征。*Journal of Big Data 7*（1）, 52。

Chiu, Y.-C., S. Zheng, L.-J. Wang, B. S. Iskra, M. K. Rao, P. J. Houghton, Y. Huang 和 Y. Chen（2021）。利用深度学习预测并表征肿瘤的癌症依赖图谱。*Science Advances 7*(34), eabh1275。

Clemmensen, L., T. Hastie, D. Witten 和 B. Ersbøll（2011）。稀疏判别分析。*Technometrics 53*(4), 406–413。

Danaher, P., P. Wang 和 D. M. Witten（2014）。用于多个类别间逆协方差估计的联合图形套索。*Journal of the Royal Statistical Society:Series B (Statistical Methodology) 76*(2), 373–397。

DepMap, Broad（2023）。Depmap 23q2 public。Figshare。数据集。

Friedman, J. H. (1989). 正则化判别分析. 美国统计学会杂志 *84*(405), 165–175.

Fukunaga, K. (1990). 统计模式识别导论 (第2版). 学术出版社.

Gaynanova, I.、J. G. Booth 和 M. T. Wells（2016）。p ¿¿ n 情形下典型向量的同时稀疏估计。*Journal of the American Statistical Associa-tion 111*（514），696–706。

Gaynanova, I. 和 M. Kolar（2015）。多组稀疏判别分析中的最优变量选择。电子统计学期刊 *9*（2），2007–2034。

Gaynanova, I. 和 T. Wang（2019）。通过线性降维实现的稀疏二次分类规则。多元分析期刊 *169*，278–299。

Golub, G. H. 和 C. F. Van Loan（2013）。矩阵计算 （第4版）。约翰斯·霍普金斯大学出版社。

Hammerman, P. S.、M. L. Sos、A. H. Ramos、C. Xu、A. Dutt、W. Zhou、L. E. Brace、B. A. Woods、W. Lin、J. Zhang 等（2011）。DDR2激酶基因中的突变确定了肺鳞状细胞癌中的一种新型治疗靶点。《癌症发现》*1*（1），78–89。

Killian, T. 和 L. Gatto (2021)。利用 DepMap R软件包处理 DepMap癌症依赖性数据。*F1000Research 10*, 416。

Kolar, M. 和 H. Liu (2014)。高维判别分析中的最优特征选择。*IEEE*信息论汇刊 *61*(2), 1063–1083。

Lˆe Cao, K.-A.、S. Boitard 和 P. Besse (2011)。稀疏偏最小二乘判别分析：面向多类别问题的生物学相关特征选择与图形展示。*BMC*生物信息学 *12*(1), 1–17。

Mai, Q., Y. Yang, 和 H. Zou（2019年）。多类别稀疏判别分析。《华人统计学》*29*(1), 97–111。

Mai, Q., H. Zou, 和 M. Yuan（2012年）。超高维稀疏判别分析的直接方法。*Biometrika 99*(1), 29–42。

Obozinski, G.、M. J. Wainwright 和 M. I. Jordan（2011）。高维多元回归中的支撑集并集恢复。《统计年刊》*39*(1)，1–47。

Ramey, J. A.、C. K. Stein、P. D. Young 和 D. M. Young（2016）。高维正则化判别分析。*arXiv* 预印本 *arXiv:1602.01182*。

Rao, C. R.（1948）。多重测量在生物分类问题中的应用。英国皇家统计学会学报：*B*辑（方法学）*10*(2)，159–203。

Sato, M.、Y. Mori、A. Sakurada、S. Fujimura 和 A. Horii（1998）。H型钙黏蛋白（cdh13）基因在人类肺癌中失活。《人类遗传学》*103*，96–101。

Shen, H., M. He, R. Lin, M. Zhan, S. Xu, X. Huang, C. Xu, W. Chen, Y. Yao, M. Mo-han 等（2019年）。PLEK2通过EGFR/CCL2通路促进胆囊癌的侵袭和转移。《实验与临床癌症研究》*38*，1–14。

Tong, T., L. Chen 和 H. Zhao（2012）。改进的均值估计及其在对角判别分析中的应用。《生物信息学》*28*(4)，531–537。

Tseng, P.（2001）。非光滑最小化的块坐标下降法的收敛性。《优化理论与应用杂志》*109*，475–494。

Wang, H.-j., Y. Gao, L. Chen, Y.-l. Li 和 C.-l. Jiang（2015）。RAB34是胶质瘤进展及预后相关生物标志物。《肿瘤生物学》*36*，1573–1578。

Witten, D. M. 和 R. Tibshirani（2011）。使用费希尔线性判别进行惩罚分类。皇家统计学会期刊：*B*辑（统计方法论）*73*(5)。

Wu, R. 和 N. Hao（2022）。通过投影进行二次判别分析。多元分析期刊 *190*, 104987。

Xu, P.、G. N. Brock 和 R. S. Parrish（2009）。用于高维微阵列数据分类的改进线性判别分析方法。计算统计与数据分析 *53*(5)，1674–1687。

Zhao, P. 和 B. Yu（2006）。论套索回归的模型选择一致性。机器学习研究杂志 *7*, 2541–2563。

Zhao, W.、A. Krishnaswamy、R. Chellappa、D. L. Swets 和 J. Weng（1998）。用于人脸识别的主成分判别分析。载于人脸识别：从理论到应用，第 73–85 页。施普林格。