



从 PCA 到 LDA：Fisher 线性判别分析的原理、算法流程、改进与应用 (Linear Discriminant Analysis, LDA)

Zijun Mao (毛子鋆) 学号: 2025439109

天津大学福州国际联合学院, 天津 300072

The International Joint Institute of Tianjin University, Fuzhou, Tianjin University, Tianjin 300072, China

zijunmao@tju.edu.cn

Abstract

在与 PCA 同步学习的课程语境下, PCA 通过方差最大化获得更优的数据表示, 而 LDA 则利用标签信息构造更具判别性的子空间。本文围绕 Fisher 于 1936 年提出的线性判别思想展开。首先, 从二分类的 Fisher 判别准则出发推导其闭式解, 并将其推广至多分类情形下的广义特征值问题, 给出清晰的算法流程与实现要点。其次, 阐释 LDA 与共享协方差的高斯生成式分类模型之间的等价关系。随后, 本文通过代表性文献梳理 LDA 在高维小样本、非线性可分、类内多模态以及可解释特征选择等场景中的典型扩展路径, 包括正则化判别分析、核化 Fisher 判别、子空间方法与小样本解法、局部 Fisher 判别, 以及稀疏/惩罚 LDA, 并补充近期鲁棒 LDA 与深度表示学习中 LDA 判别准则的当代用法。最后, 本文设计构造性示例(详见附录)说明: 当总体最大方差方向主要由无关扰动主导时, PCA 的一维投影可能退化为近似随机分类; 而 LDA 的一维投影仍可实现高精度分离, 从而直观展示 LDA 的优势及其适用边界。作业相关代码见: <https://github.com/dogmao/zijun-ml-course-tlm-ttl-quantile-sel>。

1. 引言: 从表示学习到判别学习的动机

PCA 的目标是表示。PCA 在无监督设定下寻找投影方向, 使投影后的总体方差最大, 从而获得更紧凑的低维表示。其经典工作可追溯至 Pearson (1901) 与 Hotelling (1933), 并在统计学与机器学习中形成了成熟体系 (Pearson, 1901; Hotelling, 1933; Jolliffe, 2002)。

LDA 的目标是判别。当类别标签可用时, 降维不仅用于压缩与可视化, 更服务于分类与判别。Fisher 于 1936 年提出以特征的线性组合实现类间区分, 强调应同时扩大类间差异并压缩类内散布 (Fisher, 1936)。在现代表述中, LDA 既可视为监督式降维方法, 也可视为线

性分类的统计构造; 其结果通常体现为在低维子空间中更易线性可分 (Bishop, 2006; Hastie et al., 2009)。

课程语境下的对比与理解。PCA 的投影不依赖标签 y , 主要由总体二阶统计量决定; LDA 则显式利用 y , 通过类内散布与类间散布的比值刻画判别性, 从而更直接地服务于分类任务 (Bishop, 2006)。从机器学习课程主线看, 这一差异可概括为学习目标决定表示结构: PCA 更贴近“压缩/重构”导向的表示学习, 而 LDA 更贴近“经验风险最小化”导向的判别学习。也就是说, 在容量受限(例如必须降到 1 维)时, 应优先保留对任务损失最敏感的信息, 而非仅追求总体方差最大。

2. Fisher LDA: 原理、推导与统计解释

2.1. 二分类: Fisher 判别准则与闭式解

设样本 $x \in \mathbb{R}^d$, 二分类标签 $y \in \{1, 2\}$ 。将样本投影到一维: $z = w^\top x$ 。记两类均值为 μ_1, μ_2 , 类内散布矩阵定义为

$$S_w = \sum_{i:y_i=1} (x_i - \mu_1)(x_i - \mu_1)^\top + \sum_{i:y_i=2} (x_i - \mu_2)(x_i - \mu_2)^\top. \quad (1)$$

Fisher 判别准则通过最大化投影后类间均值差与类内方差之比来确定投影方向 (Fisher, 1936; Bishop, 2006):

$$J(w) = \frac{(w^\top (\mu_1 - \mu_2))^2}{w^\top S_w w}. \quad (2)$$

对 (2) 求极值可得比例意义上的最优方向, 具有闭式形式:

$$w^* \propto S_w^{-1}(\mu_1 - \mu_2). \quad (3)$$

该表达式揭示了 LDA 的“统计矫正”含义: $(\mu_1 - \mu_2)$ 表征类中心差异, 而 S_w^{-1} 用于抑制类内相关噪声, 并在特征存在相关性时提供合理加权, 从而提升判别投影的稳定性 (Bishop, 2006; Hastie et al., 2009)。

如图 1 所示, Fisher 在原始工作中将鸢尾花的多维测量线性组合为一维判别得分, 并展示了三类样本在该得分上的频数分布 (Fisher, 1936)。可以看到, 一维线

性投影已能使 *Iris setosa* 与其余类别几乎完全分离，而 *Iris versicolor* 与 *Iris virginica* 仍存在一定重叠。这既验证了线性判别“投影后比较类条件分布”的直观含义，也提示当类别可分性不足或类内结构更复杂时，需要更高维的判别子空间，或采用核化、局部判别等扩展方法。

2.2. 多分类：散布矩阵与广义特征值问题

当类别数为 $C > 2$ 时，LDA 通常把 d 维降至 $k \leq C - 1$ 维。记总体均值 μ ，第 c 类样本数 n_c 、均值 μ_c 。定义类内散布与类间散布为

$$S_w = \sum_{c=1}^C \sum_{i:y_i=c} (x_i - \mu_c)(x_i - \mu_c)^\top, \quad (4)$$

$$S_b = \sum_{c=1}^C n_c(\mu_c - \mu)(\mu_c - \mu)^\top. \quad (5)$$

常用的多分类 Fisher 准则可写为 (Bishop, 2006; Hastie et al., 2009):

$$J(W) = \text{tr}((W^\top S_w W)^{-1}(W^\top S_b W)), \quad (6)$$

其求解等价于广义特征值问题

$$S_b v = \lambda S_w v. \quad (7)$$

取最大的 k 个特征值对应的特征向量组成投影矩阵 $W = [v_1, \dots, v_k]$ 。由于 S_b 的秩至多为 $C - 1$ ，因此 LDA 的判别子空间维度上限为 $C - 1$ (Bishop, 2006)。

2.3. 生成式视角：与共享协方差高斯模型的关系

在经典统计建模中，若各类满足高斯分布且共享协方差矩阵 Σ :

$$p(x | y = c) = \mathcal{N}(x | \mu_c, \Sigma), \quad p(y = c) = \pi_c, \quad (8)$$

则可导出线性判别函数 (Hastie et al., 2009; Bishop, 2006):

$$\begin{aligned} \delta_c(x) &= x^\top \Sigma^{-1} \mu_c - \frac{1}{2} \mu_c^\top \Sigma^{-1} \mu_c + \log \pi_c, \\ \hat{y} &= \arg \max_c \delta_c(x). \end{aligned} \quad (9)$$

该视角解释了 LDA 中“linear”的含义，并将 Fisher 判别与概率生成式分类联系起来：当共享协方差假设近似成立时，LDA 往往给出强而可解释的线性边界；当假设明显偏离（例如显著异方差或类内强非线性结构）时，则需要更一般的扩展方法 (Hastie et al., 2009)。在课程框架下，这也对应于“生成式建模假设”与“判别性能”之间的权衡：生成式假设带来可解释性与校准性，但也可能引入模型偏差。

3. 算法实现、方法扩展与应用谱系

3.1. 算法流程与实现要点

Algorithm 1 Fisher LDA: 监督式降维与判别

- 1: **Input:** 训练数据 $\{(x_i, y_i)\}_{i=1}^n$ ，类别数 C ，目标维度 $k \leq C - 1$
- 2: 计算每类均值 μ_c 与总体均值 μ
- 3: 计算类内散布 $S_w = \sum_c \sum_{i:y_i=c} (x_i - \mu_c)(x_i - \mu_c)^\top$
- 4: 计算类间散布 $S_b = \sum_c n_c(\mu_c - \mu)(\mu_c - \mu)^\top$
- 5: 求解广义特征值问题 $S_b v = \lambda S_w v$
- 6: 取最大的 k 个特征值对应向量组成投影矩阵 $W = [v_1, \dots, v_k]$
- 7: 将样本投影到判别子空间: $z = W^\top x$
- 8: 在 z 上训练或执行线性分类器（如最近均值、逻辑回归、线性 SVM）
- 9: **Output:** 投影矩阵 W 与分类器

数值稳定性与小样本问题。在高维小样本设定下， S_w 往往可能奇异，从而导致数值不稳定。实践中常用两类策略：其一引入岭正则化，令 $S_w \leftarrow S_w + \epsilon I$ ；其二先用 PCA 将维度降至不超过样本数，再执行 LDA。后者在人脸识别的 Fisherfaces 方法中尤为典型：先用 PCA 缓解维度灾难并抑制噪声，再用 LDA 强化判别结构 (Belhumeur et al., 1997; Friedman, 1989; Chen et al., 2000)。

3.2. 典型扩展：从统计稳健性到非线性与高维稀疏

正则化判别分析 (RDA)。当共享协方差假设过强，或样本不足导致协方差估计不稳定时，Friedman 提出 RDA 在 LDA 与 QDA 之间进行连续折中，并通过收缩 (shrinkage) 提升估计稳定性 (Friedman, 1989)。

核化 Fisher 判别 (KFD)。针对非线性可分情形，KFD 通过核技巧将 Fisher 判别推广到隐式高维特征空间，从而实现非线性判别投影 (Mika et al., 1999)。

小样本与零空间思想。在人脸识别等高维任务中， S_w 奇异往往是常态。Chen 等提出利用 S_w 的零空间构造判别方向，从而有效缓解 small sample size 问题 (Chen et al., 2000)。

局部 Fisher 判别 (LFDA)。当类内分布呈现多模态结构时，标准 LDA 基于全局二阶统计的建模可能不足以保留局部几何结构。Sugiyama 提出 LFDA，通过局部权重在判别性与结构保持之间取得兼顾 (Sugiyama, 2007)。

稀疏与惩罚 LDA。在 $p \gg n$ 的高维场景中，引入惩罚项可促使判别方向稀疏化，并提升可解释性。Witten 与 Tibshirani 给出惩罚 LDA 的系统构造与分析 (Witten & Tibshirani, 2011)。

异方差判别分析 (HDA)。当各类协方差显著不同且需要更一般的统计建模时，HDA 提供更灵活的判别与

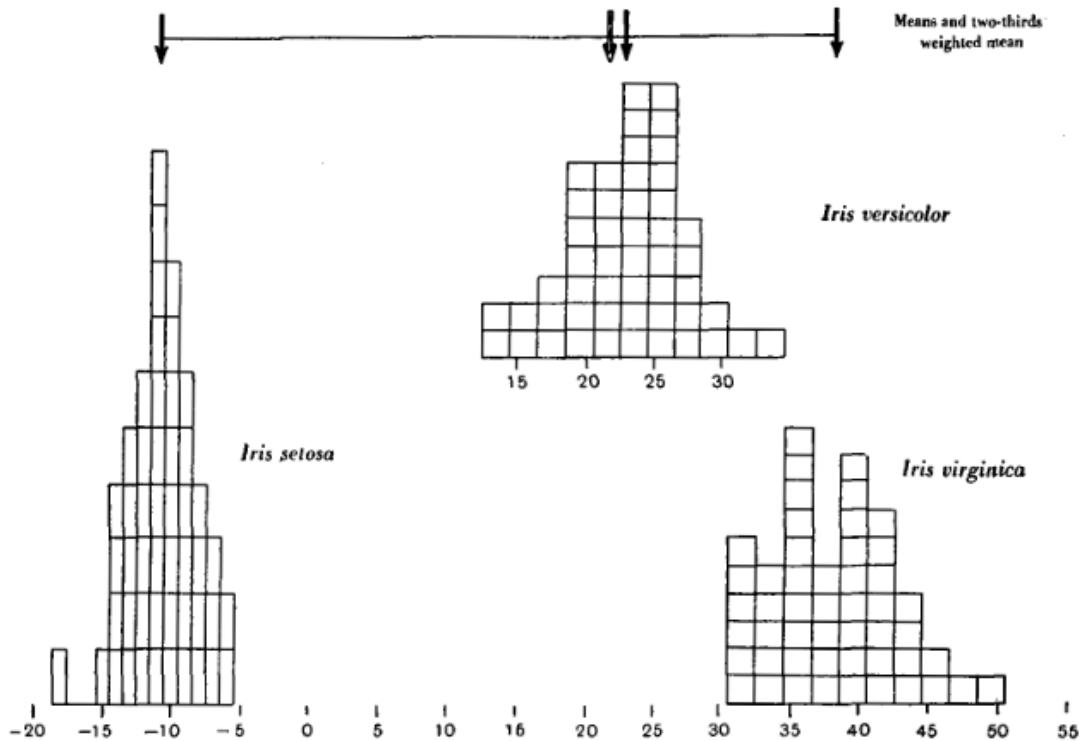


Figure 1. Fisher (1936) 在鸢尾花三类别数据上给出的判别线性函数 (discriminating linear function) 频数直方图：将多维测量线性组合为一维判别得分后，不同类别在该得分轴上呈现明显分离 (*setosa* 与其余类别几乎完全分开，而 *versicolor* 与 *virginica* 仍存在一定重叠)，直观体现了 Fisher 判别准则“增大类间差异、减小类内散布”的作用。图源：Fisher (1936) (Fig. 1)。

降维框架，并在语音识别等任务中得到应用 (Kumar & Andreou, 1998)。

近期鲁棒/高维方向的持续推进。真实数据中常见噪声污染与统计量估计不稳，尤其是均值与协方差容易受离群点与长尾分布影响。Lam 等提出基于修剪与差分分布建模的鲁棒 LDA 框架，以降低异常样本对统计量的干扰 (Lam et al., 2024)。此外，面向高维数据的稀疏化与稳健化策略仍在发展。例如，Zhuang 与 Feng 提出由空间符号 (spatial sign) 驱动的稀疏 LDA，试图在高维设定下进一步提升鲁棒性与可解释性 (Zhuang & Feng, 2026)。这些工作共同体现了课程中的方法论要点：统计估计的稳健性与判别目标的一致性同等重要。

3.3. 应用谱系：经典应用与当代表示学习连接

经典应用：监督式降维、可视化与识别。当类别数较小且需要可解释的判别投影时，LDA 常用于将数据降至 $C-1$ 维，并最大化类别分离 (Bishop, 2006)。在人脸识别中，Fisherfaces 采用“PCA 预降维 + LDA 判别投影”的组合，以缓解小样本问题，并增强对光照变化等因素的稳定性 (Belhumeur et al., 1997)；后续工作进一步从小样本处理与子空间选择等角度提升了 LDA 的可用性 (Chen et al., 2000)。在语音等任务中，异方差情形与更灵活的判别建模也被系统探讨 (Kumar & Andreou,

1998)。

当代用法：与深度表征学习的对接。在深度学习体系中，LDA 常以“判别准则”或“概率头部”的形式被复用。一方面，它可作为深度特征的线性判别探针，或作为后处理降维步骤，用于评估表征的线性可分性；另一方面，也可将 LDA 结构嵌入端到端训练，强调潜在空间中“类内紧致、类间分离”的几何性质。Tezekbayev 等在 2026 年对 Deep LDA 进行再审视，指出纯极大似然训练可能出现病态解，并提出融合判别惩罚的损失，以兼顾生成式一致性与判别性能 (Tezekbayev et al., 2026)。从课程视角看，这可理解为将经典统计判别准则转译为可微优化目标，并在现代训练框架中重新获得可用且可解释的学习信号。

4. 结论与课程反思

结论 本文在统一框架下系统介绍了 Fisher LDA 的原理、推导与算法流程：从二分类 Fisher 准则到多分类的广义特征值问题，并从生成式视角揭示其与共享协方差高斯分类器的等价关系 (Fisher, 1936; Bishop, 2006; Hastie et al., 2009)。在方法谱系方面，本文进一步梳理了正则化、核化、小样本零空间、局部判别、稀疏惩罚与异方差建模等典型扩展 (Friedman, 1989; Mika et al., 1999; Chen et al., 2000; Sugiyama, 2007; Witten &

Tibshirani, 2011; Kumar & Andreou, 1998), 并补充近期的鲁棒化与高维稀疏趋势 (Lam et al., 2024; Zhuang & Feng, 2026), 以及与深度表示学习相连接的当代用法 (Tezekbayev et al., 2026)。

构造性例子 为直观对比 PCA 与 LDA 的目标差异, 本文构造了一个二维数据示例, 使“最大方差方向”与“判别方向”不一致, 并引入仅发生在无关扰动维度上的协变量漂移 (covariate shift)。实验的可复现脚本、图像、指标与解释均置于附录。结果表明, PCA 的一维表示可能被无关的高方差扰动主导, 从而导致下游分类性能退化至近似随机; 而 LDA 的一维表示仍能对齐类间差异并保持较高的判别精度。这一现象直接对应课程中反复强调的原则: 优化目标与评估目标必须对齐——仅最大化总体方差并不保证分类风险下降。

课程反思 从机器学习课程的整体视角看, LDA 的价值不仅在于其作为经典线性分类与降维方法的地位, 更在于它提供了一种可迁移的思想模板: 当面对小样本、高维、噪声污染或分布偏移等情形时, 如何将“判别目标”(类间分离与类内紧致)与“统计估计”(均值与协方差的稳定估计)有机结合, 并在可解释性与性能之间做出理性折中。LDA 及其扩展共同勾勒出从统计学习到现代表示学习的一条清晰脉络, 有助于将课程中的经验风险最小化、偏差一方差权衡、生成式/判别式建模差异等核心概念落到具体算法之上。

致谢

感谢尚凡华老师在机器学习课程中的系统讲授与悉心指导, 使我能够在 PCA 的表示学习视角之外, 进一步理解 LDA 的判别学习思想及其在现代方法体系中的延续与发展。

References

- Belhumeur, P. N., Hespanha, J. P., and Kriegman, D. J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997. doi: 10.1109/34.598228. URL <https://doi.org/10.1109/34.598228>.
- Bishop, C. M. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York, NY, USA, 2006. ISBN 978-0-387-31073-2. URL <https://link.springer.com/book/10.1007/978-0-387-45528-0>.
- Chen, L.-F., Liao, H.-Y. M., Ko, M.-T., Lin, J.-C., and Yu, G.-J. A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, 33(10):1713–1726, 2000. doi: 10.1016/S0031-3203(99)00139-9. URL [https://doi.org/10.1016/S0031-3203\(99\)00139-9](https://doi.org/10.1016/S0031-3203(99)00139-9).
- Fisher, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936. doi: 10.1111/j.1469-1809.1936.tb02137.x. URL <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>.
- Friedman, J. H. Regularized discriminant analysis. *Journal of the American Statistical Association*, 84(405):165–175, 1989. doi: 10.1080/01621459.1989.10478752. URL <https://doi.org/10.1080/01621459.1989.10478752>.
- Hastie, T., Tibshirani, R., and Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, NY, USA, 2 edition, 2009. ISBN 978-0-387-84857-0. doi: 10.1007/978-0-387-84858-7. URL <https://doi.org/10.1007/978-0-387-84858-7>.
- Hotelling, H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 498–520, 1933. doi: 10.1037/h0071325. URL <https://doi.org/10.1037/h0071325>.
- Jolliffe, I. T. *Principal Component Analysis*. Springer Series in Statistics. Springer, New York, NY, USA, 2 edition, 2002. ISBN 978-0-387-95442-4. doi: 10.1007/b98835. URL <https://doi.org/10.1007/b98835>.
- Kumar, N. and Andreou, A. G. Heteroscedastic discriminant analysis and reduced rank HMMs for improved speech recognition. *Speech Communication*, 26(4):283–297, 1998. doi: 10.1016/S0167-6393(98)00061-2. URL [https://doi.org/10.1016/S0167-6393\(98\)00061-2](https://doi.org/10.1016/S0167-6393(98)00061-2).
- Lam, B. S., Choy, S., and Yu, C. K. Linear discriminant analysis with trimmed and difference distribution modeling. *Knowledge-Based Systems*, 299:112093, 2024. doi: 10.1016/j.knosys.2024.112093. URL <https://doi.org/10.1016/j.knosys.2024.112093>.
- Mika, S., Rätsch, G., Weston, J., Schölkopf, B., and Müller, K.-R. Fisher discriminant analysis with kernels. In *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop*, pp. 41–48. IEEE, 1999. doi: 10.1109/NNSP.1999.788121. URL <https://doi.org/10.1109/NNSP.1999.788121>.
- Pearson, K. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901. doi: 10.1080/14786440109462720. URL <https://doi.org/10.1080/14786440109462720>.

Sugiyama, M. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *Journal of Machine Learning Research*, 8:1027–1061, 2007.
URL <https://www.jmlr.org/papers/v8/sugiyama07b.html>.

Tezekbayev, M., Takhanov, R., Bolatov, A., and Assylbekov, Z. Deep linear discriminant analysis revisited, 2026. URL <https://doi.org/10.48550/arXiv.2601.01619>.

Witten, D. M. and Tibshirani, R. Penalized classification using Fisher’s linear discriminant. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(5):753–772, 2011. doi: 10.1111/j.1467-9868.2011.00783.x. URL <https://doi.org/10.1111/j.1467-9868.2011.00783.x>.

Zhuang, D. and Feng, L. Spatial sign based direct sparse linear discriminant analysis for high dimensional data. *arXiv preprint arXiv:2504.11117*, 2025.

A. 附录：可复现实验（PCA 与 LDA 在协变量偏移下的一维投影对比）

A.1. 实验目的与核心结论

本附录设计的实验，用于说明如下现象：当数据存在显著的“无关高方差方向”(nuisance high-variance direction)时，PCA 倾向于选择总体方差最大的方向作为低维表示，因此一维投影未必保留判别信息；相对地，LDA 通过最大化类间散度相对于类内散度的比值，使一维投影更直接地强化可分性，并在协变量偏移(covariate shift)情形下表现出更稳定的分类性能。

A.2. 数据构造与偏移设定

本文构造二维二分类数据 $\mathbf{x} = (x, y)$ ，其中：

- 判别信息主要位于 x 方向（两类均值在 x 方向分离），且该方向方差较小；
- y 方向为“无关扰动维度”，其方差显著更大，但两类在该维度均值相同，从而对判别无贡献；
- 测试阶段引入协变量偏移：仅增大 y 方向方差（输入边缘分布变化），而判别结构仍由 x 方向决定。

该设定体现了课程中“分布偏移下表示学习与判别学习目标可能不一致”的典型现象：最大化总体方差并不等价于最大化可分性。

A.3. 方法、指标与实现选择

投影方法。将训练集降至一维并比较：

- **PCA 1D**：无监督地选择最大方差方向；
- **LDA 1D**：监督地选择 Fisher 判别方向，最大化类间散度与类内散度之比。

判别准则 (**Fisher criterion**)。对任意方向向量 \mathbf{w} ，定义

$$J(\mathbf{w}) = \frac{\mathbf{w}^\top S_b \mathbf{w}}{\mathbf{w}^\top S_w \mathbf{w}}, \quad (10)$$

其中 S_b 为类间散度矩阵、 S_w 为类内散度矩阵。代码中分别对 PCA 主方向与 LDA 方向计算 $J(\mathbf{w})$ ，以量化“投影方向的判别性”。

一维分类器。为保持实验简洁且具有可解释性，使用 **1D 最近类均值(nearest-centroid)** 分类器：将投影后的测试点分配给距离最近的类均值。该选择避免引入复杂模型，使对比集中于“投影是否保留判别结构”。

报告指标。实验打印并对比以下量：

- 测试(shift后)准确率 acc；
- Fisher 判别准则 $J(\mathbf{w})$ ；
- PCA 第一主成分解释方差比 EVR；
- 方向夹角余弦 $\cos(\angle(\mathbf{w}_{PCA}, \mathbf{w}_{LDA}))$ ，衡量两者方向一致性。

A.4. 结果与分析（对应图与打印输出）

数值结果。在固定随机种子 (seed=0) 下，程序输出为：

Method	acc (test-shift)	$J(\mathbf{w})$	EVR	$\cos(\angle)$
PCA 1D	0.500	7.814×10^{-5}	0.749	0.0555
LDA 1D	0.963	3.557×10^{-2}	-	

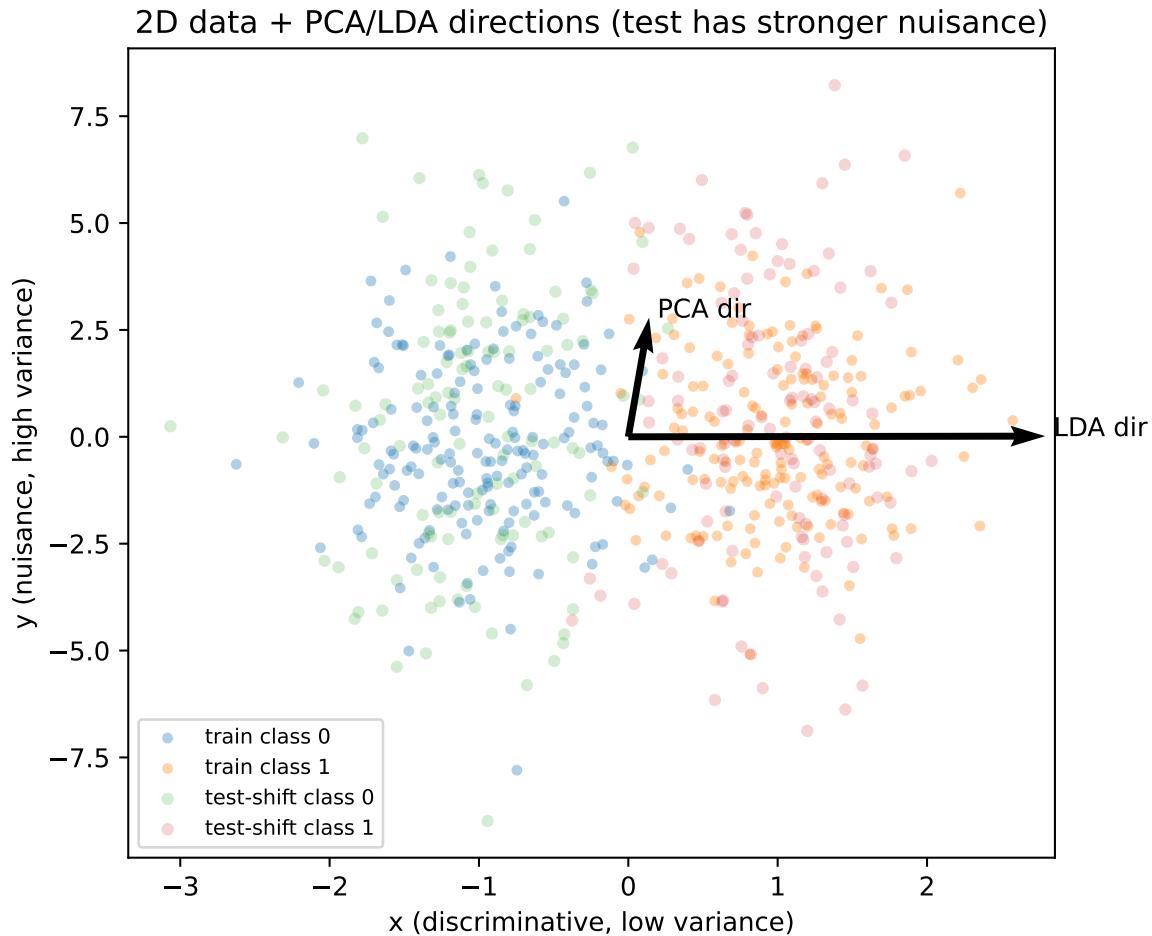


Figure 2. 训练集与协变量偏移测试集的二维散点图，并叠加 PCA 与 LDA 的一维投影方向。PCA 倾向于对齐总体最大方差的扰动维度；LDA 倾向于对齐类别可分性最强的方向。

可以观察到：尽管 PCA 的解释方差比很高 (EVR = 0.749)，其对应的一维投影并未保留有效判别信息，表现为 Fisher 判别准则极小、测试准确率接近随机。相反，LDA 所得方向的 $J(\mathbf{w})$ 显著更大，并能在协变量偏移情形下保持较高的分类准确率。两种投影方向的夹角余弦接近 0，表明 PCA 与 LDA 学得的方向几乎正交，从而揭示了“最大方差方向”与“最强判别方向”之间的结构性差异。

图 2：二维散点与投影方向的作用。图 2 在同一坐标系展示训练集与 shift 后测试集，并叠加 PCA/LDA 两个方向向量。可以看到：

- PCA 方向更接近 y 轴：该方向包含最大的总体方差，但主要来自无关扰动维度；
- LDA 方向更接近 x 轴：该方向对应类别均值差异最显著的方向，从而在一维投影中保留判别结构；
- 测试集仅在 y 维度方差增大，因此 PCA 所依赖的统计特征更易受偏移影响，而 LDA 的判别方向保持稳定。

图 3：一维投影分布的作用。图 3 展示 PCA 与 LDA 在一维投影下的类条件分布，并分别比较训练集与 shift 后测试集：

- PCA 1D 下，两类分布高度重叠，训练与测试均难以线性或阈值式分离；

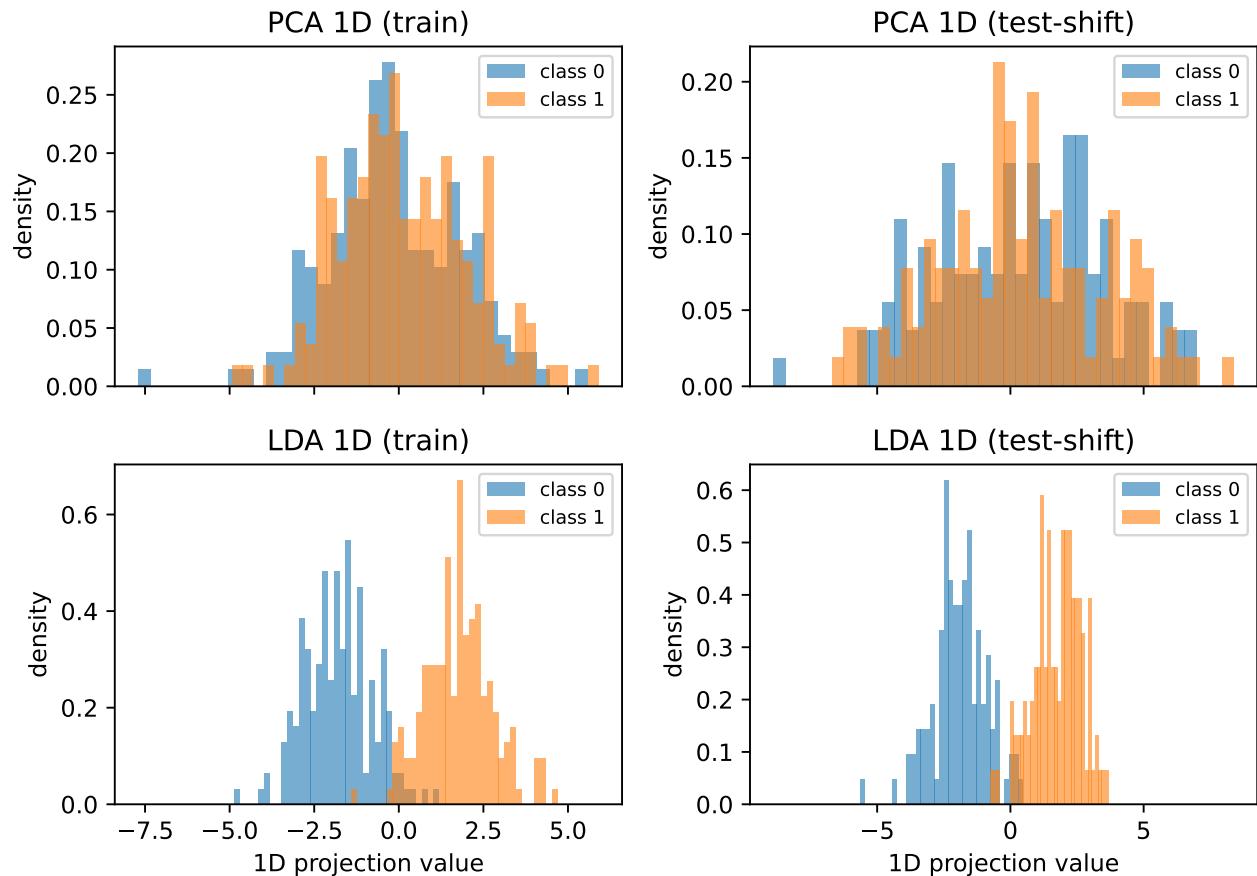


Figure 3. PCA 与 LDA 的一维投影类条件分布（训练集与协变量偏移测试集对比）。PCA 的投影导致两类高度重叠；LDA 的投影显著拉开两类间隔，从而提升一维判别性能。

- LDA 1D 下，两类分布在训练集上明显分离，并在测试偏移后仍保持分离结构，因此简单的最近均值分类即可得到高准确率。

该图直接将“表示目标（PCA）”与“判别目标（LDA）”在低维空间的差异可视化：PCA 优化总体方差，不保证类别分离；LDA 显式优化可分性，因而更契合分类任务。

A.5. 复现说明

将脚本保存为 `pca_vs_lda_shift_demo.py` 并运行：

```
python pca_vs_lda_shift_demo.py
```

程序会在终端打印上述指标，并在 `figs/` 目录下生成：`pca_lda_scatter.pdf/png` 与 `pca_lda_1d_hist.pdf/png`。在本文中，图 2 与图 3 直接由该脚本生成并插入。