Clustering Singular and Non-Singular Covariance Matrices for Classification

DATAWorks 2024

Andrew Simpson, Semhar Michael

April 17, 2024

Mathematics and Statistics Department
South Dakota State University

South Dakota State University

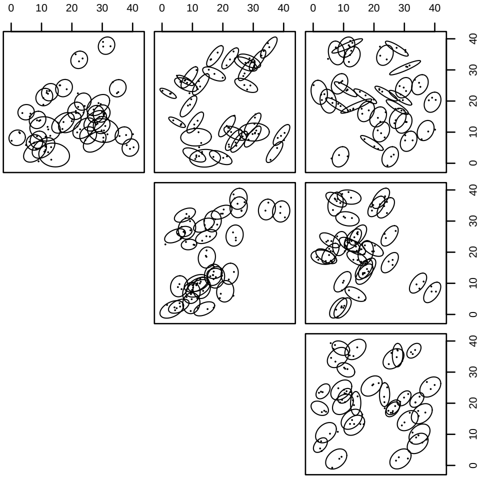
Acknowledgment

This work was partially supported by the National Science Foundation NRT: Cyber-Physical-Social System for Understanding and Thwarting the Illicit Economy (DGE 1828492, 1828302, 1828462, 1828288). The opinions and conclusions or recommendations expressed in this article are those of the authors only.

Motivating Example

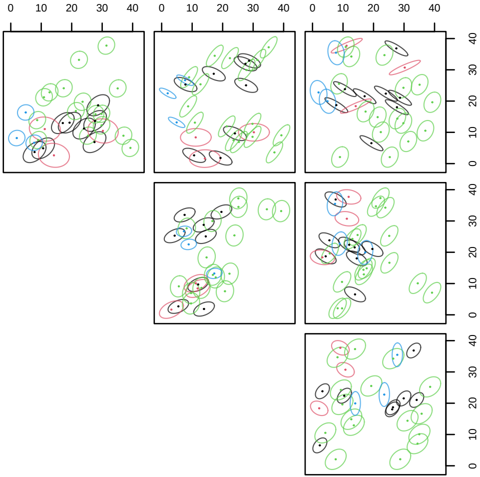
Consider a high dimensional space with with many classes and few observations per class which often appears is forensic identification of source problems.

- This example has 30 classes and 3 observations per class which follow multivariate normal distributions in a 4 dimensional space
- Quadratic Discriminant Analysis (QDA) will yield singular covariance matrices
- One would then need to use Linear Discriminant Analysis (LDA)



Motivating Example

- Assuming a shared covariance matrix is a strong assumption to make
- There may be clusters of covariance matrices shared between all classes
- **Goal:** Can we cluster common covariance matrices and pool with clusters
- This will increase the discrimination power between classes by providing stable estimates of the covariance matrices



Models

Let X_{ij} be a random vector of length p which is the *jth* observation from the *ith* class for $i \in \{1, 2, \dots, C\}$ and $j \in \{1, 2, \dots, n_i\}$.

Models

Linear Discriminant Analysis (LDA)

$$X_{ij} \sim N_p(\mu_i, \mathbf{\Sigma}),\ \hat{z} = \text{argmax}_i \phi(x; \mu_i, \mathbf{\Sigma})$$

Quadratic Discriminant Analysis (QDA)

$$X_{ij} \sim N_p(\mu_i, \mathbf{\Sigma}_i),\ \hat{z} = \text{argmax}_i \phi(x; \mu_i, \mathbf{\Sigma}_i)$$

Proposed Model: Latent Covariance Discriminant Analysis (LCDA)

$$X_{ij} | Z_i = k \sim N_p(\mu_i, \mathbf{\Sigma}_k) \text{ where } Z_i \sim \mathit{Categorical}(\pi_1, \dots \pi_K),\ \hat{z} = \text{argmax}_i \phi(x; \mu_i, \mathbf{\Sigma}_k)$$

Expectation Maximization Algorithm

The Expectation Maximization (EM) A. P. Dempster et al. algorithm is an iterative method which finds MLEs in situations where latent variables exist

E-step

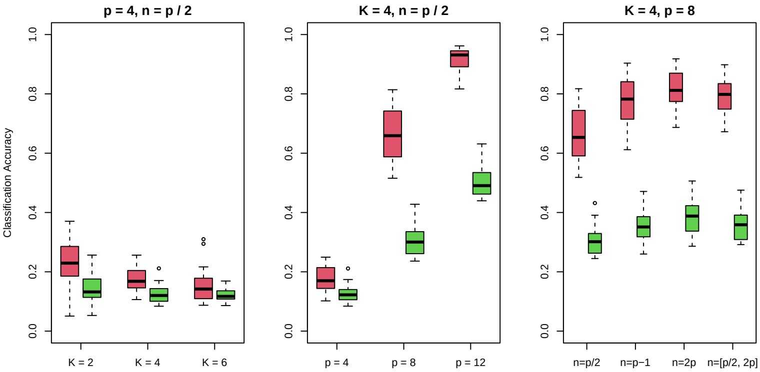
$$\tau_{ik} = \frac{\pi_k |\boldsymbol{\Sigma}_k|^{-\frac{n_i}{2}} e^{-\frac{1}{2} tr(\boldsymbol{\Sigma}_k^{-1} S_i)}}{\sum_{k'=1}^{K} \pi_{k'} |\boldsymbol{\Sigma}_{k'}|^{-\frac{n_i}{2}} e^{-\frac{1}{2} tr(\boldsymbol{\Sigma}_{k'}^{-1} S_i)}} \text{ where } S_i = \sum_{j=1}^{n_i} (x_{ij} - \mu_i)(x_{ij} - \mu_i)^T$$

M-step

$$\hat{\pi}_k = \frac{1}{C} \sum_{i=1}^{C} \tau_{ik}, \; \hat{\mu}_{\mathbf{i}} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \text{ and } \hat{\boldsymbol{\Sigma}}_k = \frac{\sum_{i=1}^{C} \tau_{ik} S_i}{\sum_{i=1}^{C} \tau_{ik} n_i}$$

Simulation Study

A stimulation study is performed in order to understand the performance of the model by looking at the out-of-sample classification rate (**LDA: Green, LCDA: Red**)

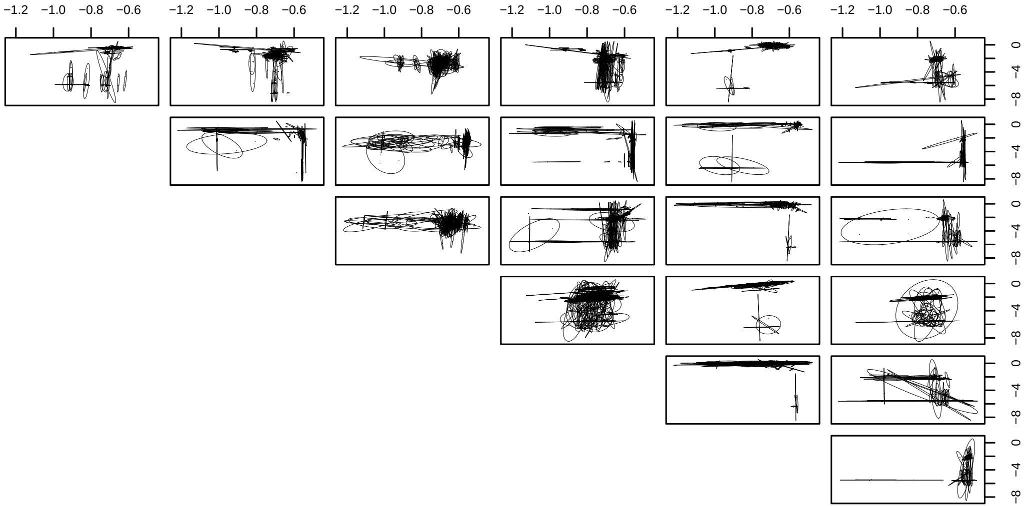


Glass Dataset Description

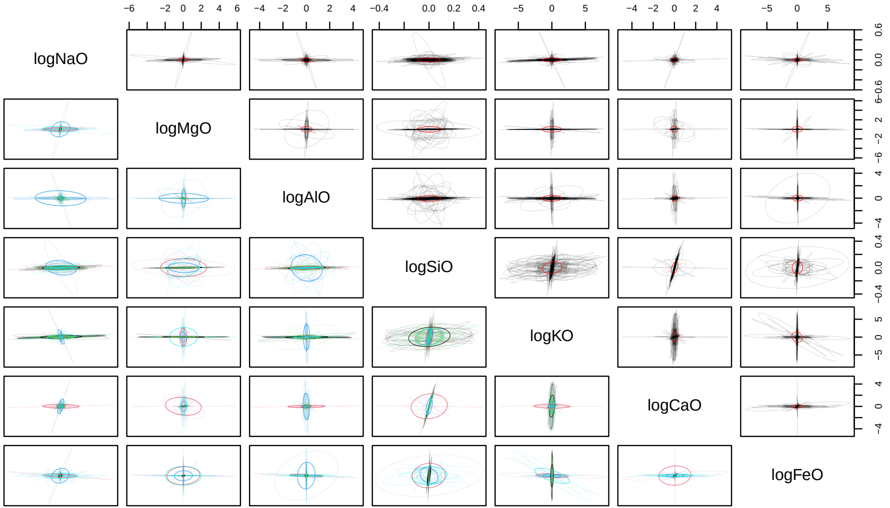
The LCDA and LDA are applied to the Zadora glass dataset Aitken et al. [2007]

- 200 Windows (classes)
- 4 fragments from each window (observations per class)
- Each measured 3 times (mean is taken)
- 7 trace elements are measured
- Ratios of Sodium, Magnesium, Aluminium, Silicon, Potassium, Calcium and Iron to Oxygen

Glass Contour Plots



Glass Results



LDA: 43%

LCDA: **57%**

Conclusion

- LDA has strong assumptions
- QDA leads to unstable or degenerate densities
- LCDA assumes there is a finite number of latent covariance matrices shared between all classes
- LCDA showed large improvements in terms of classification accuracy over LDA in situations where QDA is not even possible

References

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. 39(1):1–38. URL https://www.jstor.org/stable/2984875.

Colin G. G. Aitken, Grzegorz Zadora, and David Lucy. A Two-Level Model for Evidence Evaluation. *Journal of Forensic Sciences*, 52(2):412–419, March 2007. ISSN 0022-1198, 1556-4029. doi: 10.1111/j.1556-4029.2006.00358.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1556-4029.2006.00358.x.

José A. Díaz-García and Ramón Gutiérrez Jáimez. Proof of the conjectures of H. Uhlig on the singular multivariate beta and the Jacobian of a certain matrix transformation. *The Annals of Statistics*, 25(5), October 1997. ISSN 0090-5364. doi: 10.1214/aos/1069362383. URL https://projecteuclid.org/journals/annals-of-statistics/volume-25/issue-5/Proof-of-the-conjectures-of-H-Uhlig-on-the-singular/10.1214/aos/1069362383.full.

Sullivan Hidot and Christophe Saint-Jean. An Expectation–Maximization algorithm for the Wishart mixture model: Application to movement clustering. *Pattern Recognition Letters*, 31(14):2318–2324, October 2010. ISSN 01678655. doi: 10.1016/j.patrec.2010.07.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0167865510002217.

A.C. Rencher and W.F. Christensen. *Methods of Multivariate Analysis*. John Wiley Sons, Hoboken, 2012.

Harald Uhlig. On singular wishart and singular multivariate beta distributions. *The Annals of Statistics*, pages 395–405, 1994.

Thank you!