The mathematics of Geometric Multivariate Analysis

Stephanie Evert

7 July 2024

Contents

1 Algorithms

1.1 Linear discriminant analysis

1.1.1 The LDA algorithm

- originally proposed by Fisher (1936) for a one-dimensional discriminant between two groups
  - uses D^2/S as separation criterion where D is the difference between the group means and S the within group variance (computed from within-group covariance matrix $\mathbf{S}$)
  - directly solves for minimum, resulting in equation system $\mathbf{S}\boldsymbol{\lambda} = \mathbf{d}$
  - Fisher does not discuss an extension to multiple groups (using between-group variance as criterion) nor to a multi-dimensional discriminant
- data matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$ with n data points $\mathbf{x}_i \in \mathbb{R}^d$
- LDA algorithm as implemented in the `MASS` package is described by Venables and Ripley (2002: 331–332):
  - matrix of group means $\mathbf{M} \in \mathbb{R}^{g \times d}$ as row vectors $\mathbf{m}_j$
  - group indicator matrix $\mathbf{G} \in \mathbb{R}^{n \times g}$ with $g_{ij} = 1$ iff X_i belongs to group j
  - $\overline{\mathbf{x}} \in setR^d$ the overall mean $\overline{\mathbf{x}} = \frac{1}{n}\sum_i \mathbf{x}_i$
  - the "group predictions" are given by $\mathbf{GM}$
  - within-group covariance matrix $\mathbf{W}$ and between-group covariance matrix $\mathbf{B}$ are

$$\mathbf{W} = \frac{(\mathbf{X} - \mathbf{GM})^T(\mathbf{X} - \mathbf{GM})}{n-g}, \qquad \mathbf{B} = \frac{(\mathbf{GM} - \mathbf{1}\overline{\mathbf{x}}^T)^T(\mathbf{GM} - \mathbf{1}\overline{\mathbf{x}}^T)}{g-1} \tag{1}$$

- a one-dimensional discriminant is given by a linear combination $\mathbf{a}^T\mathbf{x}$ that maximises the ratio of between-group to within-group variance along the discriminant axis:

$$\frac{\mathbf{a}^T\mathbf{B}\mathbf{a}}{\mathbf{a}^T\mathbf{W}\mathbf{a}} \tag{2}$$

  - NB: this criterion is proportional to the F-statistic of ANOVA; since it differs only by a fixed factor, the choice of $\mathbf{a}$ also maximises the F-statistic[1]
  - to find the maximum, compute a sphering $\mathbf{y} = \mathbf{S}\mathbf{x}$ of the variables so that the within-group covariance matrix becomes $\mathbf{W}' = \mathbf{I}$
  - the problem is then to maximise $\mathbf{a}^T\mathbf{B}'\mathbf{a}$ for the transformed between-group matrix $\mathbf{B}$ subject to $\|\mathbf{a}\| = 1$ (because the transformation $\mathbf{a}' = \mathbf{S}^{-1}\mathbf{a}$ yields the same value for (2))
  - $\mathbf{a}$ is then easily found as the largest principal component of $\mathbf{B}'$
  - for an extension to a multi-dimensional discriminant, the first r principal components can be used, and the number of dimensions can be chosen according to their principal values or R^2; while this is plausible in the sphered coordinates, Venables & Ripley don't explain what separation criterion it optimises in the original coordinate system
- a different explanation of the LDA algorithm is given by Bishop (2006: 186–190), who explicitly discusses the extension to multiple classes and a multi-dimensional discriminant (Bishop 2006: 191–192)
- Bishop also points out the problem that it is no longer clear which separation criterion should be maximised and refers to Fukunaga (1990: 445–459) for a detailed exposition of different criteria and their optimisation

Useful Wikipedia articles

- Analysis of variance: `https://en.wikipedia.org/wiki/Analysis_of_variance`
- F-test: `https://en.wikipedia.org/wiki/F-test#Formula_and_calculation`
- F-distribution: `https://en.wikipedia.org/wiki/F-distribution#Definition`
- MANOVA separation criteria: `https://en.wikipedia.org/wiki/Multivariate_analysis_of_variance#Hypothesis_Testing`
- Linear discriminant analysis: `https://en.wikipedia.org/wiki/Linear_discriminant_analysis`, esp. `https://en.wikipedia.org/wiki/Linear_discriminant_analysis#Multiclass_LDA`
- Blessing of dimensionality: `https://en.wikipedia.org/wiki/Curse_of_dimensionality#Blessing_of_dimensionality` (but more relevant for Azuma paper)

1.1.2 Repeated-measures LDA

- **repeated-measures** as appropriate terminology: `https://en.wikipedia.org/wiki/Repeated_measures_design`

[1] See Wikipedia article on Analysis of variance for the usual form of the F-statistic. See Wikipedia articles on the F-test and the F-distribution for an explanation of the scaling factors involved.

2

2.1

3

3.1

References

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188.

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*. Morgan Kaufmann, San Francisco, 2nd edition.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S-PLUS*. Springer, New York, 4th edition.

Todo list