A Compact Guide to LDA (CDT-53)

Preprint · February 2021

DOI: 10.13140/RG.2.2.35120.07684/3

1 author:

Luciano da F. Costa
University of São Paulo
734 PUBLICATIONS **13,303** CITATIONS

SEE PROFILE

A Compact Guide to LDA (CDT-53)

Luciano da Fontoura Costa
luciano@ifsc.usp.br

São Carlos Institute of Physics – DFCM/USP

2nd Feb. 2021

Abstract

While the Karhunen-Loève transform, and respectively associated principal component analysis methodologies, provide optimal data transformation in the sense of concentrating variance along the first axes in unsupervised manner, it is also possible to perform statistical transformations that best separate, according to a specified criterion, a group of samples with respective categories. One example of such approaches is the so-called linear discriminant analysis – LDA. In this work, we briefly introduce the LDA approach, including two application examples.

"The sunset line hourglasses the day."

LdFC.

1 Introduction

The attribution of categories to objects is an intrinsic human activity from which ultimately language and science developed [1]. Though nothing comes with labels attached in the natural work, as soon as something become of particular relevance to humans, a category and name is soon respectively assigned.

As a consequence, an ever growing set of data with associated labels are produced as we keep on separating objects into meaningful categories while considering respective measurements of features. Welcome to the enticing area of *pattern recognition* (e.g. [2, 3, 4]), a branch of machine learning and artificial intelligence!

Having addressed the Karhunen-Loève transform, and respectively associated principal component analysis (PCA) in a previous work [5, 6], in this work we turn our attention to a closely related method, known as *linear discriminant analysis* — LDA, a *supervised* method capable of rotating, possibly in combination with dimensionality reduction, the feature space associated to a set of categorized data so as to maximize the overall separation between the involved classes according to a scatter measurements (e.g. [2, 3]).

Though LDA presupposes data normality and identical covariance matrix for each involved category, this method may still provide interesting results when the former condition is relaxed. As for the second condition, application of the standardization procedure as a preliminary step may also contribute to obtaining interesting results by making, in some situations, the involved covariance matrix more similar one another.

This work starts by presenting the total, inter- and intra-class matrices of a set of labelled measurements, and how a measurement of the separation of the groups can be obtained from these matrices. The basic steps in the LDA are then presented, followed by some examples of respective applications.

2 Data Scattering

Let the categorized data of interest be organized into a table as illustrated in Table 2.

The number of categories is henceforth represented as C, while the number of samples in each of these classes c is indicated as N_c.

In order to minimize the effect of different magnitudes of the involved features, it is important to perform standardization of the original data, which can be done by applying:

$$\hat{X}_i = \frac{X_i - \mu_{X_i}}{\sigma_{X_i}} \tag{1}$$

where μ_{X_i} and σ_{X_i} correspond to the mean and standard deviation of feature (or measurement) X_i taken along the whole data set (i.e. irrespectively of the categories).

Table 1: A typical representation of the categorized data to be analyzed: the N rows correspond to samples, and the M columns to respective measurements. A total of C categories are involved, being also indicated in the table.

Category	*Sample*	X_1	X_2	...	X_M
1	1	2.389	1001.3	...	0.0023
1	2	-4.764	818.0	...	0.0011
2	3	-1.764	208.0	...	0.0089
2	4	-1.009	111.3	...	0.0089
	...	...	...	...	...
C	N-1	6.782	183.9	...	0.0097
C	N	8.490	230.7	...	0.0097

A possible manner to represent the overall scattering of this data, irrespectively of categories and in a similar way as in PCA, is in terms of the respective *total scattering matrix* S:

$$S[i,j] = \sum_{k=1}^{N} (X_{i,k} - \mu_{X_i})(X_{j,k} - \mu_{X_j}) \tag{2}$$

where μ_{X_i} is the mean of measurement X_i, and $i, j = 1, 2, \ldots, M$.

A scattering matrix $S^{[c]}$ can be similarly defined respectively to each of the C involved classes, corresponding to the following covariance matrices:

$$S_a^{[c]}[i,j] = \sum_{k \ in \ class \ c} (X_{i,k} - \mu_{X_i}^{[c]})(X_{j,k} - \mu_{X_j}^{[c]}) \tag{3}$$

where $\mu_{X_i}^{[c]}$ is the mean of measurement X_i among the observations belonging to category c, i.e.:

$$\mu_{X_i}^{[c]} = \frac{1}{N_c} \sum_{k \ in \ class \ c} X_{i,k} \tag{4}$$

The *intra-class* scattering matrix S_A of the original data can now be defined as:

$$S_A = S_a^{[1]} + S_a^{[2]} + \ldots + S_a^{[C]} \tag{5}$$

The *inter-class* scattering matrix S_E corresponds to:

$$S_E[i,j] = \sum_{k=1}^{C} N_c(\mu_{X_i}^{[k]} - \mu_{X_i})(\mu_{X_j}^{[k]} - \mu_{X_j}) \tag{6}$$

Observe that all the matrices above have dimension $C \times C$. Thus, it is possible to combine them. For instance, it can be shown that:

$$S = S_A + S_E \tag{7}$$

3 Scattering Maximization Criterion

Let's now define the following matrix:

$$\Pi = S_A \left(S_E\right)^{-1} \tag{8}$$

which also has dimension $X \times X$.

An index quantifying the overall separation between all involved clusters corresponds to:

$$\rho = trace\left\{S_A \left(S_E\right)^{-1}\right\} \tag{9}$$

Given that ρ quantifies the overall separation between the involved categories, it is now possible to find the rotation of the data coordinate axes that maximizes ρ. This is precisely what LDA does.

It should be kept in mind that ρ assumes that the data is normal and the covariance matrix of each category are mutually identical. However, good results may still be obtained for approximations of these conditions.

4 Linear Discriminant Analysis

We are now in the position of defining more objectively the important method known as LDA.

Given a set of N observations of normally (or nearly normally) distributed data involving M measurements, and so that the covariance matrices for each category are identical (or similar), the basic steps in LDA are:

- Standardize the data;
- Calculate S_a and S_e as described in Section 2;
- Obtain Π from those two matrices, which plays a role analogous to that of the covariance matrix in PCA;
- Estimate the eigenvalues λ_i (in decreasing order), as well as the respective eigenvectors $\vec{v}_i$, of Π;
- Transform the data as in Equation 10, in a manner directly analogous to PCA.

$$\tilde{\vec{X}} = \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \vdots \\ \tilde{X}_M \end{bmatrix} = \begin{bmatrix} \leftarrow & \vec{v}_1 & \rightarrow \\ \leftarrow & \vec{v}_2 & \rightarrow \\ \vdots & \vdots & \vdots \\ \leftarrow & \vec{v}_M & \rightarrow \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix} = \\ = V[X_1 \ X_2 \ \ldots \ X_M]^T = V\vec{X} \tag{10}$$

LDA can be understood to rotate the coordinate axes so as to maximize the respective value of ρ, therefore obtaining a potentially good separation between the points

belonging to each category. It is also possible to reduce the dimension of the data representation by taking only the initial components $\tilde{X_1}, \tilde{X_2}, \ldots, \tilde{X_Q}$, with $Q < M$, in a manner similar to the principle of PCA.

It is important to keep in mind that a rotation of coordinates found to maximize the separation of a specific set of data may not necessarily work well for other sets of samples obtained for that same type of data. As always, it is always necessary to verify the obtained results in a critical manner.

5 Case Examples

In this section we present two examples of LDA application. First, we provide a simple, numeric example, involving a small number of points. Then, we present a graphical illustration of the effects of LDA considering a larger number of points. For simplicity's sake, both these examples involve two features.

Consider the original data in Table 2, with $C = 2$ categories, $N = 5$ sample points, and $M = 2$ features.

Table 2: A typical representation of the categorized data to be analyzed: the N rows correspond to samples, and the M columns to respective measurements. A total of C categories are involved, being also indicated in the table.

Category	*Sample*	X_1	X_2
1	1	10.8	10.1
1	2	10.0	10.8
1	3	10.1	10.2
2	4	0.015	0.020
2	5	0.012	0.097

First, the data is standardized, yielding the results shown in Table 3.

The scattering matrices can be obtained as:

$$S = \begin{bmatrix} 4.0000 & 3.9822... \\ 3.9822... & 4.0000 \end{bmatrix}$$

$$S_a = \begin{bmatrix} 0.01193... & -0.00721... \\ -0.00721... & 0.00906 \end{bmatrix}$$

$$S_e = \begin{bmatrix} 3.9880... & 3.9894... \\ 3.9894... & 3.9909... \end{bmatrix}$$

From which we get:

$$\Pi = \begin{bmatrix} 1156.688... & 1360.698... \\ 1157.104... & 1361.188... \end{bmatrix}$$

Table 3: Standardization of the original data, as a preparation to LDA application.

Category	*Sample*	$\hat{X}_1$	$\hat{X}_2$
1	1	0.817...	0.682...
1	2	0.676...	0.806...
1	3	0.693...	0.699...
2	4	-1.093...	-1.101...
2	5	-1.094...	-1.087...

The LDA matrix therefore is:

$$V = \begin{bmatrix} -0.7069... & -0.7619... \\ -0.7072... & 0.6476... \end{bmatrix}$$

with $\rho = 2517.875....$

Table 4 presents the new variables obtained after LDA.

Table 4: The result of LDA.

Category	*Sample*	$\tilde{X}_1$	$\tilde{X}_2$
1	1	-1.098...	-0.1364...
1	2	-1.092...	0.0439...
1	3	-1.023...	-0.0372...
2	4	1.161...	0.060...
2	5	1.601...	0.069...

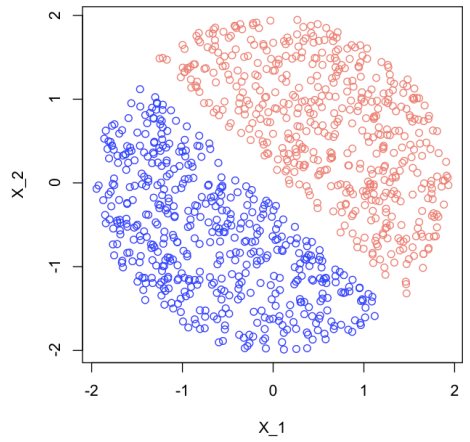
Our second example involves the set of points in Figure 1(a), which have $N = 1000$ points characterized by $M = 2$ features, divided into two categories ($C = 2$).

The result of the application of LDA is shown in Figure 1(b), corresponding to a rotation of the coordinate system so as to optimize the separation of the involved categories according to a linear discriminant.

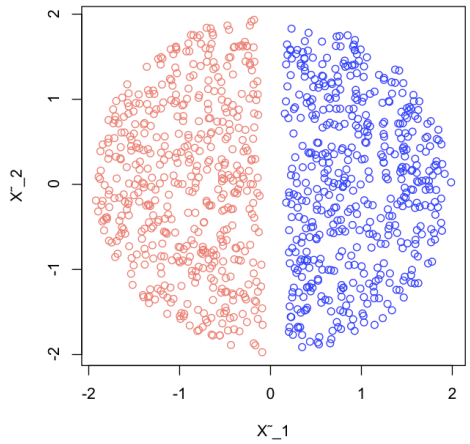
6 Concluding Remarks

As a consequence of the almost unlimited potential for applications, allied to the several involved challenges, data analysis and pattern recognition represent a critical continuing issue in science and technology.

One of the challenges while searching for clusters or heterogeneities in data distributions as defined in respective



(a)



(b)

Figure 1: Original data (a) and respectively obtained LDA (b).

feature spaces concerns the identification of separations between two or more categories of entities.

In case these separations have a linear or nearly linear nature (hence the term *linear* in LDA), and provided the data distribution follows normal distribution and that identical covariance matrices are observed for each category, it is possible to apply LDA in order to identify the most prominent separations between the involved categories.

In this work, we developed a synthetic presentation of the basic methodology known as linear discriminant analysis, starting with the presentation of the involved scattering matrices and proceeding to the characterization of data separation in terms of the product of the inter-class by the inverse of the intra-class scattering matrices. The so-obtained matrix can then be used in a manner analogous to the covariance matrix in principal component analysis (PCA), implementing a rotation of the coordinate axes so that the separation of the categories is maximized along the first axes.

A simple numeric example involving a small number of points, as well a graphic illustration of LDA over a larger number of elements, were then presented.

Acknowledgments.

Luciano da F. Costa thanks Éverton F. da Cunha for commenting on this work and to CNPq (grant no. 307085/2018-0) and FAPESP (grant 15/22308-2) for support.

References

[1] L. da F. Costa. Modeling: The human approach to science. Researchgate, 2019. https://www.researchgate.net/publication/333389500_Modeling_The_Human_Approach_to_Science_CDT-8. [Online; accessed 1-Oct-2020.].

[2] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley Interscience, 2000.

[3] K. Koutrombas and S. Theodoridis. *Pattern Recognition*. Academic Press, 2008.

[4] L. da F. Costa. Pattern cognition, pattern recognition. Researchgate, Dec 2019. https://www.researchgate.net/publication/338168835_Pattern_Cognition_Pattern_Recognition_CDT-19. [Online; accessed 29-Feb-2020].

[5] L da F. Costa. A compact guide to lda. Researchgate, 2020. https://www.researchgate.net/publication/346656784_A_Compact_Guide_to_PCA_CDT-47. accessed 10-Dez-2020.

[6] F. Gewers, G. R. Ferreira, H. F. Arruda, F. N. Silva, C. H. Comin, D. R. Amancio, and L. da F. Costa. Principal component analysis: A natural approach to data exploration. Researchgate, 2019. https://www.researchgate.net/publication/324454887_Principal_Component_Analysis_A_Natural_Approach_to_Data_Exploration. accessed 1-Oct-2020.

Costa's Didactic Texts – CDTs

CDTs intend to be a halfway point between a formal scientific article and a dissemination text in the sense that they: (i) explain and illustrate concepts in a more informal, graphical and accessible way than the typical scientific article; and (ii) provide more in-depth mathematical developments than a more traditional dissemination work.

It is hoped that CDTs can also incorporate new insights and analogies concerning the reported concepts and methods. We hope these characteristics will contribute to making CDTs interesting both to beginners as well as to more senior researchers.

Each CDT focuses on a limited set of interrelated concepts. Though attempting to be relatively self-contained, CDTs also aim at being relatively short. Links to related material are provided in order to provide some complementation of the covered subjects.

Observe that CDTs, which come with absolutely no warranty, are non distributable and for non-commercial use only.

Please check for new versions of CDTs, as they can be revised. Also, CDTs can and have been cited, e.g. by including the respective DOI. Please cite this CDT in case you use it, so that it may also be useful to other people. The complete set of CDTs can be found at: `https://www.researchgate.net/project/Costas-Didactic-Texts-CDTs`, and a respective guide at: `https://www.researchgate.net/publication/348193269_A_Guide_to_the_CDTs_CDT-0`