

Project 2 - Fisherwalks: Statistical gait recognition

Amélia O. F. da S.
190037971@unb.br

Departamento de Ciência da
Computação
Universidade de Brasília
Campus Darcy Ribeiro, Asa Norte
Brasília-DF, CEP 70910-900, Brazil

Abstract

Fisher’s Linear Discriminant Analysis is a commonly used tool for dimensionality reduction and linear classification. This paper proposes a new application of Fisher’s Discriminant to distinguish between subjects using gait features, facilitating subject identification when other methods such as facial identification are unfeasible.

1 Introduction

When two or more populations have been measured in several characters x_1, \dots, x_8 , special interest attaches to certain linear functions of the measurements by which the populations are best discriminated.

Ronald A. Fisher - "The Statistical Utilization Of Multiple Measurements"
1936

Fisher’s Linear Discriminant Analysis (henceforth “LDA”) has been utilised in a multitude of fields across its long history to classify and differentiate populations, from medical data analysis to face recognition (where it is widely known as the Fisherfaces method).

LDA fundamentally provides us with a projection in which dispersion within the classes is minimised and class separation is maximised[1]. This allows a multi-variable problem to be treated as a single-variable problem or a simpler multi-variable problem by projecting the data onto axes which describe a more marked distinction between the classes[2][3].

This paper proposes an application for LDA as a way to make subject recognition simpler and applicable on a wider range of contexts through identification by gait analysis. The proposed method allows for identification in situations where facial recognition and other methods might fail or where features might be intentionally concealed.

2 Methodology

2.1 Pose analysis

The *OpenPose*[4][5][6][7] system was used to provide a rough estimate of the subjects’ 3D poses (precise positioning would require multiple views or depth cameras) at four different

key points in their walk cycle: the points of maximum and minimum leg extension for both left and right paces.

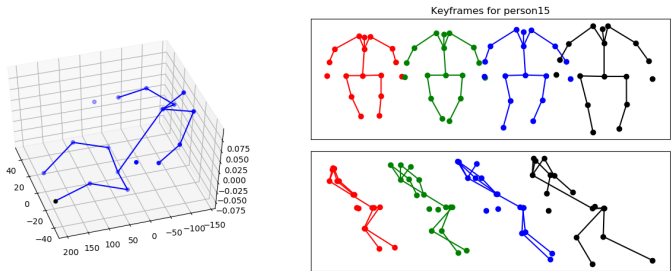


Figure 1: Sample output for OpenPose and a full collected sample

From each pose estimate, eight angle measurements were made:

- Left and right torso-arm angles ("armpit")
- Left and right elbow angles
- Left and right hip-thigh angles
- Left and right knee angles

That provides eight scale-independent features for each point in time, resulting in 32 angle dimensions. The standard deviation for each key angle was also calculated and added as a feature for a description of their variation along the walk cycle. This set of features describes a 40-dimensional sample vector.

2.2 Linear Discriminant Analysis

For any given set of samples for two or more classes, there exists a projection axis (which can also be seen as a function of the original axes) on which the projections of the classes will show maximum class mean separation and minimal spread within the classes[1]. Assuming the classes are unimodal and follow a Gaussian distribution, we can calculate that axis using Fisher's Linear Discriminant.

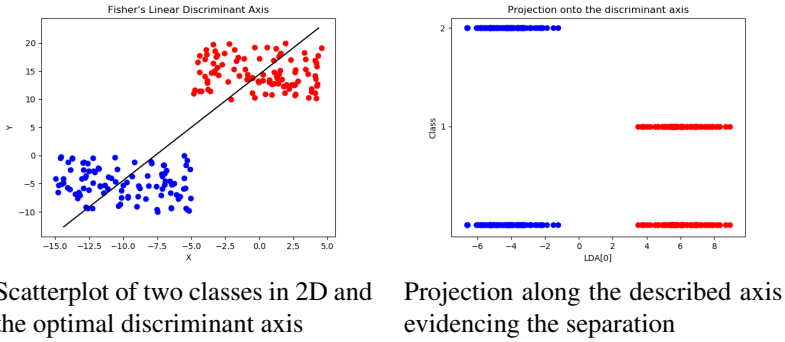


Figure 2: A 2D visualisation of Fisher's Linear Discriminant

Fisher's linear discriminant can be calculated using *Scatter Matrices*, which approximate the covariance between variables in a set[8]. Since the final goal is to maximise the scatter *between* classes and minimise the scatter *within* the classes, two scatter matrices will be needed:

$$S_B := N_{classes} * Scatter(class\ means), S_W := \sum_{Classes} Scatter(samples_{class}) \quad (1)$$

With

$$Scatter(X) := \sum_{i=0}^{i=N_X} (vector_i - Mean(X))(vector_i - Mean(X))^T \quad (2)$$

The simultaneous maximisation of scatter "between" classes and minimisation of scatter "within" classes can be represented as the maximisation of a ratio between both (see Appendix 1.1 [8] for a thorough description of that maximisation process). The vector W that maximises that ratio will then be defined as:

$$W := argmax_W (|W^T S_B S_W^{-1} W|) \quad (3)$$

Solving that maximisation problem is equivalent[8] to solving an eigenvector problem for that resulting matrix, allowing us to easily compute the optimal vectors.

2.3 One-against-all projections

LDA allows us to calculate the optimal $N - 1$ axes to separate N classes, but for multi-class classification problems we can also adopt a more intuitive (and often more effective) "One-against-all" approach, also known as "Binary Relevance"[8]. That approach consists in focusing on individually separating classes from all others instead of generating a general classifier. For that, multiple discriminant analyses were used to find N axes that individually separate each class from the others, greatly simplifying classification.

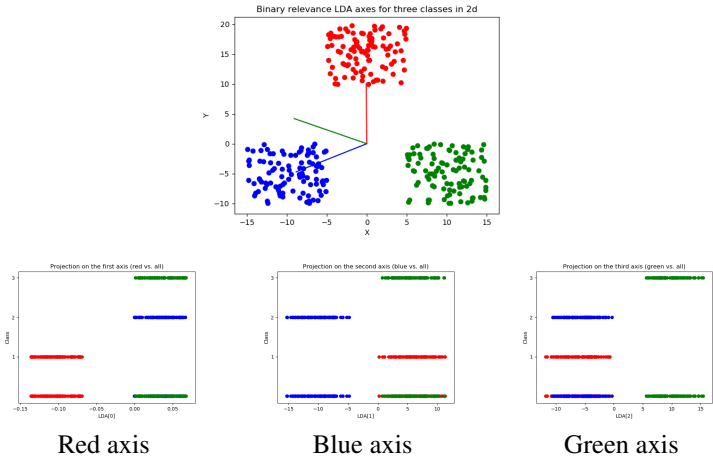
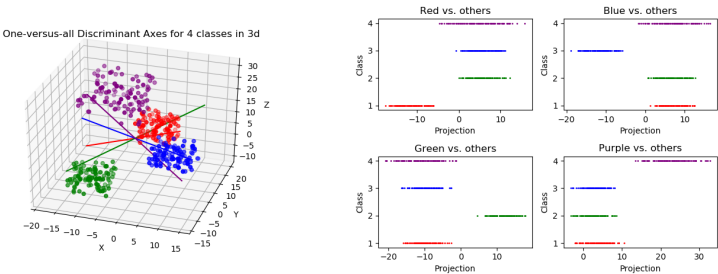


Figure 3: 2D One-against-all discriminant axes for 3 classes



Scatterplot of four classes in 3D and the "One-versus-all" axes Projections along the resulting axes evidencing the separation between each class and the others

Figure 4: 3D One-against-all discriminant axes for 4 classes

2.4 Principal Component Analysis

Calculating the scatter matrices (eq. 1) yields a problem when dealing with many dimensions: those matrices will be singular when the number of dimensions is greater than the number of samples used, turning the inverse matrix calculation impossible[9].

A solution for this problem is to first apply other dimensionality reduction methods to the data so the number of dimensions matches the number of samples[9]. Here, *Principal Component Analysis* was used to extract the $\min(N_{class})$ projection axes along which the data has the most variance.

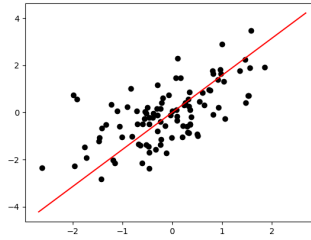


Figure 5: A 2D visualisation of the the first axis generated by PCA

The Principal Component Analysis problem is mathematically equivalent to the "least squares" problem or Eignevector decomposition and can be performed as following:

Given a set S of vectors (our full dataset)

$$W := \operatorname{argmin}_W (|\operatorname{Scatter}(S)|) \quad (4)$$

Which is also an eigenvector problem (see eq. 3 and Appendix 1.1 [1]).

3 Results

3.1 Implementation description

The statistical methods above were implemented using the Numpy[1] scientific computing package in Python.

Due to processing power limitations, the key images for each sample had to be manually separated and then processed through the proposed pipeline. This limited the test sample size to four classes of six 4-image samples each, totalling 96 images.

Provided with the extracted sample vectors, the program generates a single projection matrix P that combines the PCA and LDA projections as following:

Let D be a data matrix containing n 40-dimensional column vectors of c classes and a minimum number m of samples per class.

Let R be our resulting projected data matrix.

$$\begin{aligned} R_{c \times n} &:= LDA_{c \times m} PCA_{m \times 40} D_{40 \times n} \\ R_{c \times n} &= (LDA_{c \times m} PCA_{m \times 40}) D_{40 \times n} \\ R_{c \times n} &= P_{c \times 40} D_{40 \times n} \end{aligned} \quad (5)$$

For ease of implementation, line vector notation was used in the code instead of the standard column vector notation, meaning in practice calculations will be done in the reverse order (right to left, with all matrices transposed).

3.2 Feature importances

It is possible to estimate the importance of each measured feature by measuring its final weight on the projection[6].

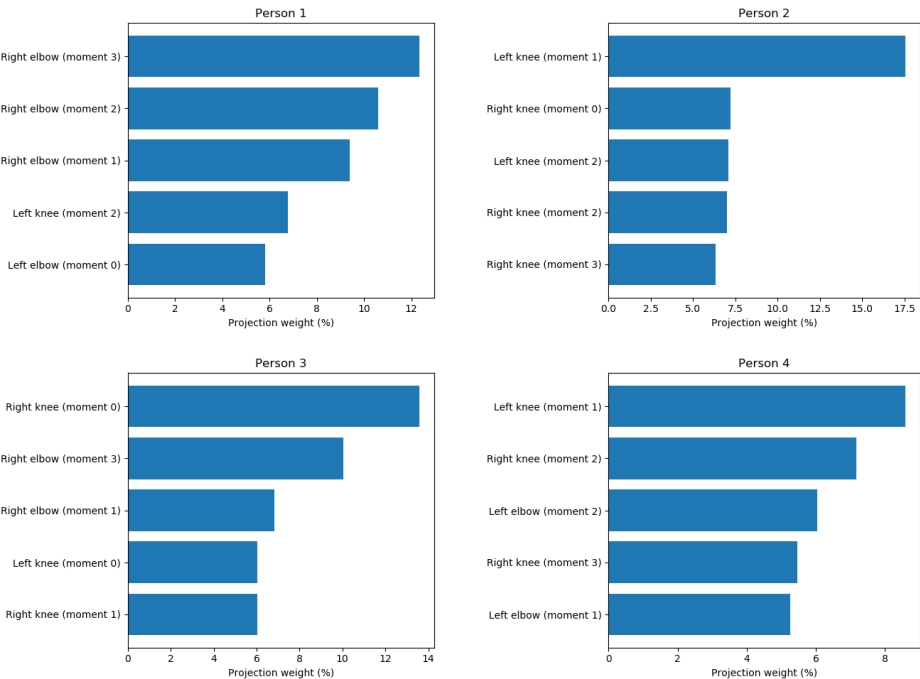


Figure 6: Feature importances for the four subjects

Those results can be interpreted as the main discriminating characteristics for each sub-
ject[9]. Those characteristics reflect visually perceptible differences between the subjects
(See Appendix 1.4 [9]).

3.3 Classification tests

Due to the small number of samples, the tests were limited to evaluating how well the pro-
jections describe the class space instead of the typical "training" vs. "testing" sets. The
following methods were used for that purpose:

3.3.1 Nearest Mean

Given the projection axes, each sample's label was assigned according to the closest class
mean on that projection (Euclidean distance). All samples were assigned the correct class.

The LDA projection can significantly improve nearest mean classification on situations
where axes are imbalanced scale-wise[9] (see Appendix 1.2 [9]).

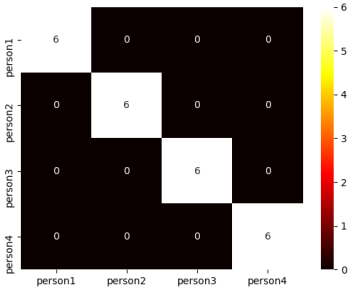


Figure 7: Nearest Mean Confusion Matrix

3.3.2 Thresholding on individual dimensions

An advantage of using Binary Relevance classification is the possibility of establishing a single threshold ($\frac{\mu_A + \mu_b}{2}$) on each axis for deciding whether a sample belongs to a class (linear classification)[13]. This allows for quickly determining whether individuals belong to a certain class without the need to compare them with other references.

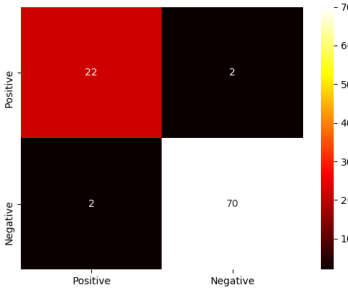


Figure 8: Single-Axis thresholding Confusion Matrix

This method yielded 4 misclassifications out of 96 samples in total (see Appendix 1.3 [14] for more information on what caused those misclassifications).

4 Conclusion

Although the number of samples was limited by the processing power available, the results obtained seem to prove the method’s usefulness in practical applications. The resulting projections not only accurately separated the classes but also matched visually perceptible and intuitive patterns (see Appendix 1.1 [15]).

The proposed method would benefit from further research, especially on automatically extracting the key images used in processing (through faster pose extraction methods) and testing the resulting performance on larger datasets.

References

- [1] Authors. Appendix 1, 2021. Supplied as additional material appendix.pdf. 322
- [2] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 323
- [3] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*, 2017. 324
- [4] R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936. 325
- [5] R. A. Fisher. The statistical utilization of multiple measurements. *Annals of Eugenics*, 8(4):376–386, 1938. 326
- [6] W. Härdle and L. Simar. Applied multivariate statistical analysis. 2003. 327
- [7] Charles R. Harris, K. Jarrod Millman, St’efan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fern’andez del R’io, Mark Wiebe, Pearu Peterson, Pierre G’erard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL <https://doi.org/10.1038/s41586-020-2649-2>. 328
- [8] João P Hespanha Peter N. Belhumeur and David Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711–720, 1997. 329
- [9] Sarunas Raudys and Anil Jain. Small sample size effects in statistical pattern recognition: Recommendations for practitioners. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 13:252–264, 04 1991. 330
- [10] Tomas Simon, Hanbyul Joo, Iain Matthews, and Yaser Sheikh. Hand keypoint detection in single images using multiview bootstrapping. In *CVPR*, 2017. 331
- [11] Laurens Van Der Maaten, Eric Postma, and Jaap Van den Herik. Dimensionality reduction: a comparative review. *J Mach Learn Res*, 10:66–71, 2009. 332
- [12] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Convolutional pose machines. In *CVPR*, 2016. 333
- [13] Min-Ling Zhang, Yu-Kun Li, Xu-Ying Liu, and Xin Geng. Binary relevance for multi-label learning: an overview. *Frontiers of Computer Science*, 12(2):191–202, 2018. 334