Journal of Machine Learning Research 9 (2008) 2489-2490 Submitted 4/08; Revised 6/08; Published 11/08

On the Equivalence of Linear Dimensionality-Reducing Transformations

Marco Loog[*] M.LOOG@TUDELFT.NL
ICT Group
Delft University of Technology
Mekelweg 4
2628 CD Delft, The Netherlands

Editor: Leslie Pack Kaelbling

Abstract

In this JMLR volume, Ye (2008) demonstrates the essential equivalence of two sets of solutions to a generalized Fisher criterion used for linear dimensionality reduction (see Ye, 2005; Loog, 2007). Here, I point out the basic flaw in this new contribution.

Keywords: linear discriminant analysis, equivalence relation, linear subspaces, Bayes error

1. Introduction

Some time ago, Ye (2005) studied an optimization criterion for linear dimensionality reduction and tried to characterize the family of solutions to this objective function. The description, however, merely covers a part of the full solution set and is therefore, in fact, not at all a characterization. Loog (2007) has corrected this mistake, giving the proper, larger set of solutions. In this volume, Ye (2008) now demonstrates that the two solution sets are essentially equivalent.

In principle, Ye (2008) is correct and the two sets of dimension reducing transforms can indeed be considered equivalent. At the base of this fact is that mathematically speaking anything can be equivalent to anything else. The point I would like to convey, however, is that the equivalence considered is not essential and, as a result, the two sets are in fact essentially different. The main question in this is what is 'essential' in the context of supervised linear dimensionality reduction?

2. Essential Equivalence

Concerned with classification tasks, the performance of every dimensionality reduction criterion should primarily be discussed in relation to the Bayes error (see Fukunaga, 1990, Chapter 10). As such, transformations might be considered essentially equivalent if their Bayes errors in the reduces spaces are equal. A closely related definition is to consider transformations A and B equivalent if there is a nonsingular transformation T such that $A = T \circ B$ (see Fukunaga, 1990). The latter is more restrictive than the former as the existence of T implies an equal Bayes error for A and B, but the implication in the other direction does not necessarily hold. When A and B are linear and there is such a transform T, both of them span the same subspace of the original feature space, obviously

∗. Also in the Image Group, University of Copenhagen, Universitetsparken 1, 2100 Copenhagen Ø, Denmark.

resulting in the equality of the Bayes errors. Based on the foregoing, two linear transformations are also considered essentially equivalent if they span the same subspace.

Now, without providing any rationale, Ye (2008) declares two linear transformations A and B to be equivalent if there is a vector v such that $A(x_i - v) = B(x_i - v)$ for all feature vectors x_i in the training set. The following very simple examples demonstrate, however, why the latter definition of equivalence is flawed.

Let $x_1 = (0,0)^t$ and $x_2 = (1,0)^t$ be two training samples, $A = (1,0)$, $B = (-1,0)$, $C = (1,1)$, $D = (0,0)$, and $E = (0,1)$ be linear transformations, and let v equal to $(v_1, v_2)^t$. Now, firstly, one cannot have both $-v_1 = A(x_1 - v) = B(x_1 - v) = v_1$ and $1 - v_1 = A(x_2 - v) = B(x_2 - v) = -1 + v_1$, and therefore A is not equivalent to B even thought $A = -B$. That is, two transforms that trivially define the same subspace are apparently not equivalent. Secondly, $D(x_i - v) = 0 = E(x_i - v)$ shows that transforms spanning subspaces of different dimensions can be equivalent. Finally, a straightforward calculation shows that A and C are equivalent, that is, two transforms that obviously span different subspaces, and therefore most probably result in different Bayes errors, are considered equivalent.

3. In Conclusion

Maintaining that the equivalence relation in Ye (2008) is flawed, it directly follows that it cannot be concluded that the different sets of solutions as given by Loog (2007) and Ye (2005) are essentially equivalent. In fact, as should be obvious from Loog (2007), they are essentially different. Given that x_1 and x_2 (as defined above) come from two different classes, one can easily check that the solution set by Ye (2005) is given by $\{(a,0) | a \in \mathbb{R} \backslash 0\}$, that is, nondegenerate multiples of $A = (1,0)$, while the true set also contains transformations like $C = (1,1)$. Both define different subspaces and, generically, lead to different Bayes errors.

Acknowledgments

This research is supported by the Innovational Research Incentives Scheme of the Netherlands Research Organization [NWO], the Netherlands, and the Research Grant Program of the Faculty of Science, University of Copenhagen, Denmark.

References

K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, New York, 1990.

M. Loog. A complete characterization of a family of solutions to a generalized fisher criterion. *Journal of Machine Learning Research*, 8:2121–2123, 2007.

J. Ye. Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems. *Journal of Machine Learning Research*, 6:483–502, 2005.

J. Ye. Comments on the complete characterization of a family of solutions to a generalized fisher criterion. *Journal of Machine Learning Research*, 9:517–519, 2008.