

Handwritten digit recognition with Nonlinear Fisher Discriminant Analysis

Pietro Berkes

Institute for Theoretical Biology
Humboldt University Berlin
Invalidenstr a e 43, D - 10115 Berlin, Germany
p.berkes@biologie.hu-berlin.de
<http://itb.biologie.hu-berlin.de/~berkes>

Abstract. To generalize the Fisher Discriminant Analysis (FDA) algorithm to the case of discriminant functions belonging to a nonlinear, finite dimensional function space \mathcal{F} (Nonlinear FDA or NFDA), it is sufficient to expand the input data by computing the output of a basis of \mathcal{F} when applied to it [1–4]. The solution to NFDA can then be found like in the linear case by solving a generalized eigenvalue problem on the between- and within-classes covariance matrices (see e.g. [5]). The goal of NFDA is to find linear projections of the expanded data (i.e., nonlinear transformations of the original data) that minimize the variance within a class and maximize the variance between different classes. Such a representation is of course ideal to perform classification. The application of NFDA to pattern recognition is particularly appealing, because for a given input signal and a fixed function space it has no parameters and it is easy to implement and apply. Moreover, given C classes only $C - 1$ projections are relevant [5]. As a consequence, the feature space is very small and the algorithm has low memory requirements and high speed during recognition.

Here we apply NFDA to a handwritten digit recognition problem using the MNIST database, a standard and freely available set of 70,000 handwritten digits (28×28 pixels large), divided into a training set (60,000 digits) and a test set (10,000 digits). Several established pattern recognition methods have been applied to this database by Le Cun et al. [6]. Their paper provides a standard reference work to benchmark new algorithms.

We perform NFDA on spaces of polynomials of a given degree d , whose corresponding basis functions include all monomials up to order d in all input variables. It is clear that the problem quickly becomes intractable because of the high memory requirements. For this reason, the input dimensionality is first reduced by principal component analysis. On the preprocessed data we then apply NFDA by expanding the training patterns in the polynomial space and solving the linear FDA eigenvalue problem. As mentioned above, since we have 10 classes we only need to compute the first 9 eigenvectors. Since the within-class variance is minimized, the patterns belonging to different classes tend to cluster in the feature space when projected on the eigenvectors. For this reason we classify the digits with a simple method such as Gaussian classifiers.

We perform simulations with polynomials of degree 2 to 5. With polynomials of degree 2 the explosion in the dimensionality of the expanded space with increasing number of input dimensions is relatively restricted, so that it is possible to use up to 140 dimensions. With higher order polynomials one has to rely on a smaller number of input dimensions, but since the function space gets larger and includes new nonlinearities, one obtains a remarkable improvement in performance. The best performance is achieved with polynomials of degree 3 and 35 input dimensions, with an error rate of 1.5% on test data. This error rate is comparable to but does not outperform that of the most elaborate algorithms (Table 1). The performance of NFDA is however remarkable considering the simplicity of the method and the fact that it has no a priori knowledge on the problem, in contrast for example to the LeNet-5 algorithm [6] which has been designed specifically for handwritten character recognition. In addition, for recognition, NFDA has to store and compute only 9 functions and has thus small memory requirements and a high recognition speed.

It is also possible to formulate NFDA using the *kernel trick*, in which case one can in principle use function spaces of infinite dimensionality [1–4]. However, the limiting factor in that formulation is the number of training patterns, which makes it not realistic for this application. The performance of NFDA could be further improved using for example a more problem-specific preprocessing of the patterns (e.g., by increasing the size of the training set with new patterns generated by artificial distortion of the original one), boosting techniques, or mixture of experts with other algorithms [5, 6].

METHOD	% ERRORS
Linear classifier	12.0
K-Nearest-Neighbors	5.0
1000 Radial Basis Functions, linear classifier	3.6
Best Back-Propagation NN (3 layers with 500 and 150 hidden units)	2.95
Reduced Set SVM (5 deg. polynomials)	1.0
LeNet-1 (16 × 16 input)	1.7
LeNet-5	0.95
Tangent Distance (16 × 16 input)	1.1
Nonlinear Fisher Discriminant Analysis (3 deg. polynomials, 35 input dim)	1.5

Table 1. Performance comparison Error rates on test data of various algorithms. All error rates are taken from [6].

Keywords: nonlinear fisher discriminant analysis, pattern recognition, digit recognition, feature extraction

This work has been supported by a grant from the Volkswagen Foundation.

References

1. Mika, S., Rätsch, G., Weston, J., Schölkopf, B., Müller, K.R.: Fisher discriminant analysis with kernels. In Hu, Y.H., Larsen, J., Wilson, E., Douglas, S., eds.: *Neural Networks for Signal Processing*. Volume IX. (1999) 41–48 *Proceedings of the IEEE Signal Processing Society Workshop*.
2. Baudat, G., Anouar, F.: Generalized discriminant analysis using a kernel approach. *Neural Computation* **12** (2000) 2385–2404
3. Mika, S., Rätsch, G., Weston, J., Schölkopf, B., Smola, A., Müller, K.R.: Invariant feature extraction and classification in kernel spaces. In Solla, S., Leen, T., Müller, K.R., eds.: *Advances in Neural Information Processing Systems*. Volume 12. (2000) 526–532
4. Mika, S., Smola, A., Schölkopf, B.: An improved training algorithm for kernel Fisher discriminants. In Jaakkola, T., Richardson, T., eds.: *Proceedings AISTATS 2001*. (2001) 98–104
5. Bishop, C.M.: *Neural Networks for Pattern Recognition*. Oxford University Press (1995)
6. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86** (1998) 2278–2324