

CONTENTS

Contents	1
1 Discriminant Component Analysis	2
1.1 Fisher's LDA	2
1.2 DCA Algorithms	2
1.3 Applications of DCA	2
1.4 DCA Algorithm Iteration Steps	2
1.5 Comparison of DCA and LDA	3
2 Experiment	4
2.1 Face Recognition	4
2.2 Synthetic data	4

1 DISCRIMINANT COMPONENT ANALYSIS

1.1 Fisher's LDA

Fisher's LDA involves solving the following problem:

$$W = \arg_T \max \frac{|T^T S_b T|}{|T^T S_w T|} \quad (1)$$

where S_b and S_w are the within-class and between class scatter matrices respectively.

And the solute onto this problem can be obtained analytically by solving a generalized eigenvalue problem:

$$S_b W = \Lambda S_w W + \Gamma W \Sigma_w^{-1} W^T S_b + \Theta, \quad (2)$$

However, there are two drawbacks with LDA based methods: (1) the number of linear bases is limited by the number of classes, and (2) the bases are not orthogonal in general.

1.2 DCA Algorithms

DCA Scheme starting from the full signal space, we find the most discriminatory (optimal) projection D1 solving (2). Then we construct a subspace which is orthogonal to that optimal projection and again find an optimal projection in the subspace. By repeating the above procedure, we obtain a new feature representation for the original signal. At least two DCA algorithms are available: the first one is suitable for any-class problems, the second one is valid only for two-class problems but more robust.

1.3 Applications of DCA

Here we propose new data-dependent distance metrics based on DCA. Initially, the weight vector (eigenvalues) $\mu = [\mu_1, \mu_2, \dots, \mu_n]$ and the linear mapping matrix $D = [D_1, D_2, \dots, D_n]$ are produced by DCA based on available samples. We then linearly map the original signal x into a new feature y :

$$y = D^T x.$$

Assuming x_1 and x_2 are two samples, and y_1, y_2 are the corresponding features in the transformed space, then one new distance metric is defined as follows::

$$d_{DCA}(y_1, y_2) = \sqrt{\sum_{i=1}^n \mu_i (y_{1,i} - y_{2,i})^2}. \quad (3)$$

It is noted that when the estimated matrices S_w and S_b are perturbed far away from the true ones or the second-order statistics is not sufficient to describe the underlying distribution, it is better not to impose the weights strictly. Thus, we propose a more general distance metric (GDCA):

$$d_{GDCA}(y_1, y_2) = \left(\sum_{i=1}^n (f(\mu_i, \alpha) + \beta) \cdot (y_{1,i} - y_{2,i})^2 \right)^{1/2}, \text{ where } f \text{ is a monotonic function of } \mu'_i, \quad (4)$$

and also a function of the dimension/size ratio $\frac{n}{N}$ and the complexity of the underlying distribution h .

1.4 DCA Algorithm Iteration Steps

- Initialization of DCA algorithm by computing mean vectors, covariance matrix, and scatter matrices;

- Iteration starts with solving an eigenproblem to find the most discriminatory projection;
- Discriminant basis and eigenvalue are obtained and a Gram-Schmidt operation is performed;
- Preparation for the next iteration involves computing a new matrix and new scatter matrices;

1.5 Comparison of DCA and LDA

Comparison:

- (1) DCA provides more discriminatory information than LDA, which makes distance metrics more robust (Figure 1);
- (2) DCA is characterized by an elliptical NN rule, unlike the rectilinear NN rule of LDA, which makes it less sensitive to changes in parameter;
- (3) Orthogonal DCA bases provide advantages in robustness of execution compared to non-perpendicular LDA bases, especially in case of estimation errors;
- (4) When changing the parameters, both methods deviate from the ideal projections, but it is assumed that the change in the percentage of the DCA distance is less than that of the LDA;
- (5) Changing the shape of the NN DCA rule presumably causes less performance degradation than that of the LDA;
- (6) Simulation on three-class data shows that the distance changes for DCA are significantly less than for LDA, on average 47.4% versus 100.6%, respectively.

2 EXPERIMENT

2.1 Face Recognition

The experiment we conducted here is the application of subspace DCA for face recognition and its comparison with a subspace LDA system. The face subspace was obtained by training on 1038 FERET images from total 444 classes with only 2 to 4 training samples per class. We perform simple normalization to all input face images and the normalized image size is chosen to be 48×42 . Finally, the near-optimal dimension of the subspace is chosen to be 300 (4). We apply both methods (subspace LDA and subspace DCA) to a probe/testing set of 115 images and gallery set of 738 images. While both methods deliver almost identical better performance using soft LDA/DCA distance metric (refer to for more detail). Strict LDA distance metric have far worse performance than strict DCA distance metric (3). When soft metrics are used, the ratio of minimal weight and maximal weight used for LDA is 0.12 while for DCA the normalized weights dropped rapidly with the 65th weight being 5.4×10^{-4} .

2.2 Synthetic data

For the synthetic data, we performed two types of experiments. The first type uses true mean vectors and covariance matrices, or those robustly estimated from a large number of samples, to carry out Linear Discriminant Analysis (LDA) and Discriminant Component Analysis (DCA). The second type estimates these parameters with a limited number of samples and then carries out LDA and DCA.

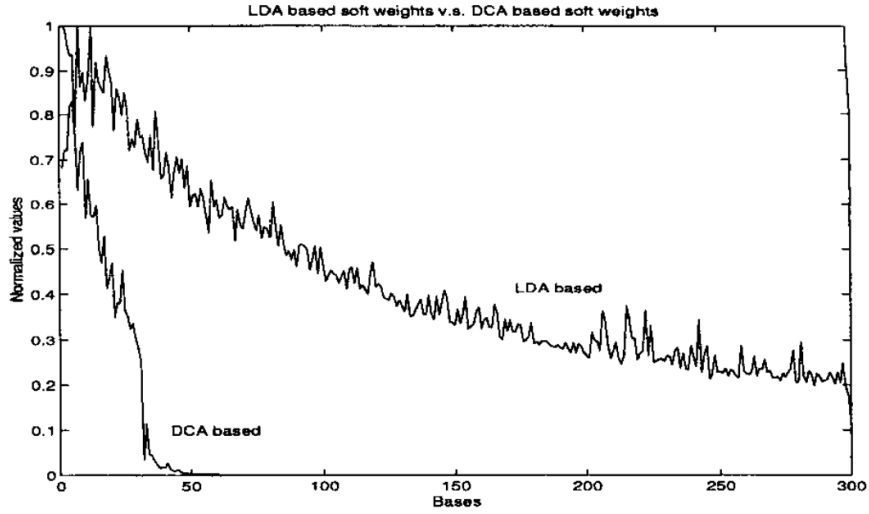


Fig. 1. Comparison of soft weights in DCA and LDA metrics

The purpose of doing this is mainly to verify the idea of DCA conceptually without worrying about the parameter estimation problem. It also casts some insight into how robust the new method is. This in these experiments we list two performance scores: the left one is for true parameters while the right one is for estimated parameters.

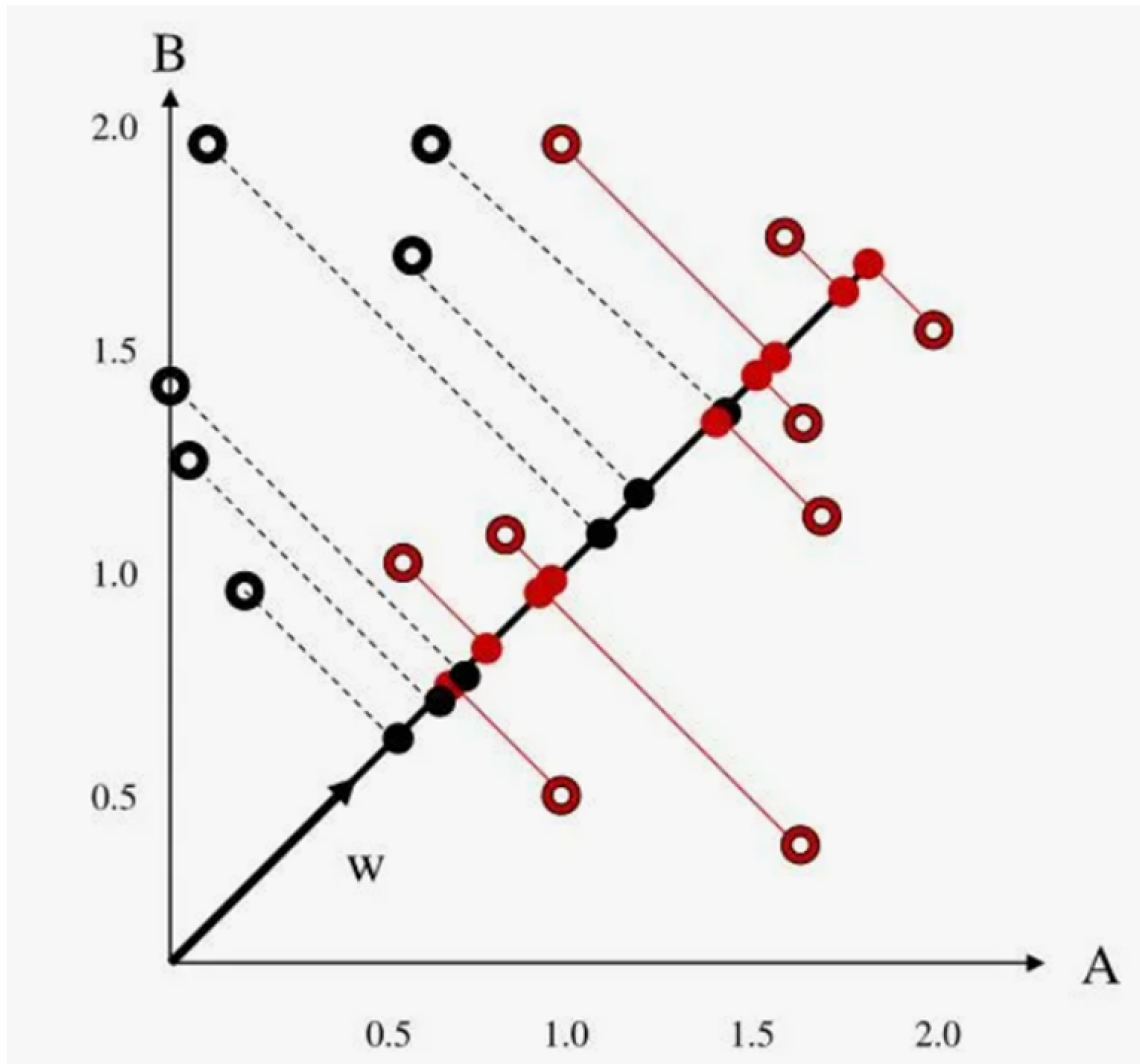


Fig. 2. Dimensionality reduction

For the synthetic data, we performed two types of experiments. The first type uses the true mean vectors and covariance matrices (or robustly estimated from large number of samples) to carry out LDA and DCA, while the second type estimates these parameters with limited number of samples and then carries out LDA and DCA.

Table 1. Performance/error comparison when the underlying problem has close ideal Bayes error and LDA error: .1465 and .167.

Classifier	LDA	LDA-NN	DCA-NN	NN
$N_i = 10$.249	.245/.270	.250/.268	.282
$N_i = 100$.180	.240/.244	.241/.241	.208

Table 2. Performance/error comparison when the underlying problem has very different ideal Bayes error and LDA error: .089 and .207.

Classifier	LDA	LDA-NN	DCA-NN	NN
$N_i = 10$.288	.296/.312	.269/.292	.273
$N_i = 100$.218	.288/.289	.235/.242	.208

The first experiment (Table 1) reports the performance comparison among different classifiers with a Gaussian distributed dataset which has close ideal Bayes and LDA errors based on true parameters. The second experiment (Table 2) reports the performance comparison among different classifiers with a Dimensionality reduction (Figure 2) which has quite different ideal Bayes and LDA errors.