Incremental and Decremental Exponential Discriminant Analysis for Face Recognition

Nitin Kumar, National Institute of Technology, Uttarakhand, India

R.K. Agrawal, School of Computer & Systems Sciences, Jawaharlal Nehru University, New Delhi, India

Ajay Jaiswal, S. S. College of Business Studies, University of Delhi, Delhi, India

ABSTRACT

Linear Discriminant Analysis (LDA) is widely used for feature extraction in face recognition but suffers from small sample size (SSS) problem in its original formulation. Exponential discriminant analysis (EDA) is one of the variants of LDA suggested recently to overcome this problem. For many real time systems, it may not be feasible to have all the data samples in advance before the actual model is developed. The new data samples may appear in chunks at different points of time. In this paper, the authors propose incremental formulation of EDA to avoid learning from scratch. The proposed incremental algorithm takes less computation time and memory. Experiments are performed on three publicly available face datasets. Experimental results demonstrate the effectiveness of the proposed incremental formulation in comparison to its batch formulation in terms of computation time and memory requirement. Also, the proposed incremental algorithms (IEDA, DEDA) outperform incremental formulation of LDA in terms of classification accuracy.

INTRODUCTION

Appearance based methods (Turk & Pentland, 1991; Murase & Nayar, 1995) are being widely used for feature extraction in the recent past for face recognition. In these methods, the facial image of a person with size \( l \times w \) pixels is represented as a vector in \( n \)-dimensional space where \( n = l \times w \). Such image data are usually characterized with high dimensions and small-sample-size (SSS). This necessitates dimensionality reduction to avoid curse-of-dimensionality (Bellman, 1961) prior to develop any learning model. This also improves performance in terms of accuracy of the learning model, computation time and memory storage. One of popular dimensionality reduction methods is a Linear Discriminant Analysis (LDA) (Fukunaga, 1990; Duda, Hart & Stork, 2000) which is a supervised technique and aims at finding an optimal transformation \( W \) that maximizes the between-class scatter and minimizes the within-class scatter simultaneously. If \( S_b \) and \( S_w \) denote the between-class and within-class scatter matrices respectively, then the Fisher’s criterion is given by (Duda, Hart & Stork, 2000):

DOI: 10.4018/ijcvip.2014010104

Keywords: Illumination, Merging, Scatter, Small Sample Size, Splitting
arg \ \max_{W} \left| \frac{W^{T}S_{b}W}{W^{T}S_{w}W} \right| \tag{1}

The optimal transformation matrix $W$ is computed by solving the following generalized eigenvalue decomposition problem:

$$S_{w}W = \lambda S_{w}W$$ \tag{2}

The resultant transformation provides a more compact representation of the original data with preservation of salient features for classification. LDA has been successfully implemented in applications in which there are enough samples to be analyzed. However, when the dimensionality of the samples is large as compared to the number of samples, LDA suffers from SSS problem. Under such situations, the \textit{within-class} scatter matrix $S_{w}$ becomes singular.

In literature, several approaches have been proposed to address the SSS problem. Raudys \& Duin (1998) suggested replacing the inverse of $S_{w}$ with its pseudoinverse. Belhumeur \textit{et al.} proposed Fisherfaces (Belhumeur, Hespanha \& Kriegman, 1997) which employs principle component analysis (PCA) (Duda, Hart \& Stork, 2000) for dimensionality reduction prior to LDA. However, due to PCA transformation, some discriminant information may get lost which is useful for classification. Chen \textit{et al.} (Chen, Liao, Ko, Lin \& Yu, 2000) proposed null-space LDA (NLDA) which modifies the Fisher criterion as:

$$\arg \ \max_{W} \left| \frac{W^{T}S_{b}W}{W^{T}S_{w}W} \right|$$ \tag{3}

The original criterion is modified in order to find the discriminative information in the null space of \textit{within-class} scatter by maximizing the \textit{between-class} scatter. In NLDA, the discriminative information present outside the null space is discarded which may be useful for classification. To exploit discriminative information in the principle space in addition to the information present in null space, Dual-space LDA is proposed by Wang and Tang (2004). However, it is pointed out by Jiang \textit{et al.} (2008) that in Dual-space LDA, features are not scaled well in the null space by the average eigenvalue of \textit{within-class} scatter matrix and proposed Eigenfeature Regularization to overcome this limitation. In their approach, the image space spanned by eigenvectors of \textit{within-class} scatter matrix is decomposed into three subspaces: (a) a reliable subspace which is spanned mainly by the variation in face (b) an unstable subspace due to noise and finite number of training samples and (c) null space. Eigenfeatures are regularized differently in these subspaces based on an eigenspectrum model. LDA via QR decomposition (LDA/QR) (Ye \& Li, 2005) is another variant of LDA which involves two stages. In the first stage, QR decomposition is applied to the \textit{between-class} scatter. In the second stage, both \textit{within-class} and \textit{between-class} scatter matrices are computed from the resultant of the first stage. The first step can be used independently for dimensionality reduction but performs poorly as it involves only the \textit{between-class} scatter matrix. This method overcomes the singularity problems of classical LDA and achieves efficiency and scalability simultaneously.

Recently Zhang \textit{et al.} (2010) proposed exponential discriminant analysis (EDA), which is a generalized discriminant analysis based on matrix exponential and modifies the Fisher’s criterion as follows (Zhang, Fang, Tang, Shang \& Xu, 2010):

$$\arg \ \max_{W} \left| \frac{W^{T} \exp(S_{b})W}{W^{T} \exp(S_{w})W} \right| \tag{4}$$

Experimental results showed that EDA outperformed LDA and its variants such as PCA+LDA (Fisherfaces) (Belhumeur, Hespanha \& Kriegman, 1997), NLDA (Chen, Liao, Ko, Lin \& Yu, 2000), LDA/QR (Ye \& Li, 2005), etc.
Related Content

Rotation Invariant Texture Image Retrieval with Orthogonal Polynomials Model
[www.irma-international.org/article/rotation-invariant-texture-image-retrieval/64184/](www.irma-international.org/article/rotation-invariant-texture-image-retrieval/64184/)

Salient Region Detection for Biometric Watermarking
[www.irma-international.org/chapter/salient-region-detection-biometric-watermarking/48322/](www.irma-international.org/chapter/salient-region-detection-biometric-watermarking/48322/)

Formalising Design Patterns as Model Transformations
[www.irma-international.org/chapter/formalising-design-patterns-model-transformations/8156/](www.irma-international.org/chapter/formalising-design-patterns-model-transformations/8156/)

Ensemble Classifier for Benign-Malignant Mass Classification
[www.irma-international.org/article/ensemble-classifier-benign-malignant-mass/78372/](www.irma-international.org/article/ensemble-classifier-benign-malignant-mass/78372/)

Edge Detection by Maximum Entropy: Application to Omnidirectional and Perspective Images
[www.irma-international.org/chapter/edge-detection-maximum-entropy/77038/](www.irma-international.org/chapter/edge-detection-maximum-entropy/77038/)