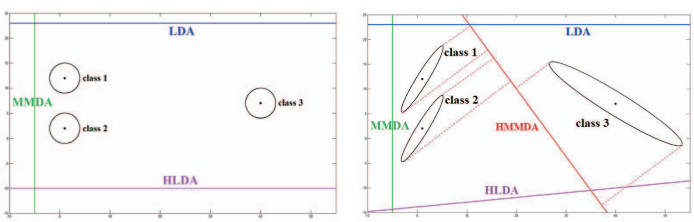
Heteroscedastic Max-min Distance Analysis

Bing Su[1], Xiaoqing Ding[1], Changsong Liu[1], Ying Wu[2]
[1]Department of Electronic Engineering, Tsinghua University. [2]Department of Electrical Engineering and Computer Science, Northwestern University.

Dimensionality reduction (DR) has become a ubiquitous procedure in many pattern recognition and machine learning applications. Among a number of DR approaches, a class of linear supervised techniques referred to as *discriminant analysis (DA)* has received a lot of attention. Linear discriminant analysis (LDA) [2] using the Fisher criterion is probably the most widely used method, which is developed under the homoscedastic Gaussian assumption. However, the covariances of different classes are often not equal in practice. Many DA approaches such as Subclass DA [7] and Marginal Fisher Analysis [6] have been proposed to attenuate this problem. Heteroscedastic LDA (HLDA) [3] explicitly models heteroscedastic data by utilizing the Chernoff criterion.

These methods actually maximize the average of all pairwise distances between classes, which will cause the so-called "class separation" problem [4]. Specifically, they tend to pay close attention to classes with larger distances, but ignore those with smaller distances, resulting in the overlap of "neighbouring" classes in the projected subspace on the basis of a specific distance measure as shown in Fig. 1(a). Max-min distance analysis (MMDA) [1] addresses this problem by maximizing the minimum pairwise distance in the latent subspace, but it is developed under the homoscedastic assumption. The heteroscedasticity of data may significantly depredates the performance of MMDA as shown in Fig. 1(b).



(a) (b)

Figure 1: (a) An illustration of the class separation problem. (b) An illustration of the influence of heteroscedastic data. The means of the three classes in (b) are the same as those in (a), while the covariances are different. MMDA can address the class separation problem as shown in (a) but fails in heteroscedastic case as shown in (b). The proposed HMMDA can successfully separate the three classes in (b).

This paper proposes Heteroscedastic MMDA (HMMDA) methods that explore the discriminative information in the difference of intra-class scatters. *Whitened HMMDA (WHMMDA)* maximizes the minimal pairwise Chenoff distance in the latent subspace following the whitening process. Under the heteroscedastic assumption, Chernoff distance d_{Cij} between classes i and j takes the discriminative information within difference of covariances into account:

$$\begin{aligned} d_{Cij} &= (\hat{\mathbf{m}}_i-\hat{\mathbf{m}}_j)^T\hat{\Sigma}_{ij}^{-1}(\hat{\mathbf{m}}_i-\hat{\mathbf{m}}_j)\\ &+\frac{1}{\alpha_{ij}(1-\alpha_{ij})}\log\frac{|\hat{\Sigma}_{ij}|}{|\hat{\Sigma}_i|^{\alpha_{ij}}|\hat{\Sigma}_j|^{1-\alpha_{ij}}}\end{aligned} \tag{1}$$

where $\alpha_{ij}=p_i/(p_i+p_j)$, $\hat{\mathbf{m}}_i$, $\hat{\Sigma}_i$ are the mean and variance of class i in the whitened space, and $\hat{\Sigma}_{ij}=\alpha_{ij}\hat{\Sigma}_i+(1-\alpha_{ij})\hat{\Sigma}_j$. It has been shown in [3] that d_{Cij} can be obtained as the trace of a positive semi-definite matrix $\mathbf{S}_{Cij}$.

$$\begin{aligned}\mathbf{S}_{Cij} &= \hat{\Sigma}_{ij}^{-1/2}(\hat{\mathbf{m}}_i-\hat{\mathbf{m}}_j)(\hat{\mathbf{m}}_i-\hat{\mathbf{m}}_j)^T\hat{\Sigma}_{ij}^{-1/2}\\ &+\frac{1}{\alpha_{ij}(1-\alpha_{ij})}(\log\hat{\Sigma}_{ij}-\alpha_{ij}\log\hat{\Sigma}_i-(1-\alpha_{ij})\log\hat{\Sigma}_j)\end{aligned} \tag{2}$$

Hence the objective of WHMMDA is:

$$\begin{aligned}&\max_{\mathbf{W}_2}\ \min_{1\le i<j\le C}\ (p_ip_j)^{-1}tr(\mathbf{W}_2{}^T\mathbf{S}_{Cij}\mathbf{W}_2)\\ &s.t.\ \ \mathbf{W}_2{}^T\mathbf{W}_2=\mathbf{I}_{d'}\end{aligned} \tag{3}$$

(3) can be optimized by introducing variable $\mathbf{Z}=\mathbf{W}\mathbf{W}^T$ and relaxing the feasible set of $\mathbf{Z}$ to its convex hull following [1].

Orthogonal HMMDA (OHMMDA) explicitly obtains a joint optimization of maximizing the minimum pairwise Chernoff distance between classes and minimizing the diffusion within the same classes with the orthogonal constraint in a trace quotient fashion:

$$\begin{aligned}&\max_{\mathbf{W}}\ \frac{\min_{1\le i<j\le C}\ (p_ip_j)^{-1}tr(\mathbf{W}^T\mathbf{S}_{OCij}\mathbf{W})}{tr(\mathbf{W}^T\mathbf{S}_w\mathbf{W})}\\ &s.t.\ \ \mathbf{W}^T\mathbf{W}=\mathbf{I}_{d'}\end{aligned} \tag{4}$$

$\tilde{\mathbf{S}}_{OCij}=(p_ip_j)^{-1}\mathbf{S}_{OCij}$. (4) can be solved by the bisection search strategy proposed in [5] by relaxing the feasible set of $\mathbf{Z}$ and deriving the upper and lower bound of the auxiliary valuable t.

Different formulations of $\mathbf{S}_{OCij}$ and $\mathbf{S}_w$ represent the dissimilarity between classes and the compactness of classes in different aspects. Encoding margin information is a commonly used strategy [5, 6]. The intra-class scatters can be used to force adjacent samples to be close.:

$$\Sigma_i^m=\sum_{\mathbf{x}_p\in class\, i}\ \sum_{\mathbf{x}_q\in N_k^i(\mathbf{x}_p)}(\mathbf{x}_p-\mathbf{x}_q)(\mathbf{x}_p-\mathbf{x}_q)^T \tag{5}$$

where $N_k^i(\mathbf{x}_p)$ denotes the set of k-nearest neighbors in class i. We propose the pairwise inter-class scatter of margin $\mathbf{S}_{Cij}^m$ to consider the difference between locally similarity within classes at the second-order level.

$$\begin{aligned}\mathbf{S}_{Cij}^m &= (\Sigma_{ij}^m)^{-1/2}(\mathbf{m}_i-\mathbf{m}_j)(\mathbf{m}_i-\mathbf{m}_j)^T(\Sigma_{ij}^m)^{-1/2}\\ &+\frac{1}{\alpha_{ij}(1-\alpha_{ij})}(\log\Sigma_{ij}^m-\alpha_{ij}\log\Sigma_i^m-(1-\alpha_{ij})\log\Sigma_j^m)\end{aligned} \tag{6}$$

where $\Sigma_{ij}^m=\alpha_{ij}\Sigma_i^m+(1-\alpha_{ij})\Sigma_j^m$. Two variants of OHMMDA, namely mar-OHMMDA and sup-OHMMDA, are developed through substituting the intra-class scatter of margin and the support pairwise inter-class scatters for the within-class and pairwise between-class scatters, respectively.

To evaluate the effectiveness of the proposed HMMDA methods, we perform experiments on five UCI datasets and the Yale Face database with three different classifiers including the nearest neighbor classifier (1-NN), the nearest mean classifier (NM) and the quadratic classifier. HMMDA algorithms achieve superior performances compared against several popular DR approaches including LDA, HLDA and MMDA.

[1] W. Bian and D. Tao. Max-min distance analysis by using sequential sdp relaxation for dimension reduction. *PAMI*, 33(5):1037–1050, 2011.

[2] R.A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.

[3] M. Loog and R.P.W. Duin. Linear dimensionality reduction via a heteroscedastic extension of lda: the chernoff criterion. *PAMI*, 26(6):732–739, 2004.

[4] Marco Loog, Robert P. W. Duin, and Reinhold Haeb-Umbach. Multiclass linear dimension reduction by weighted pairwise fisher criteria. *PAMI*, 23(7):762–766, 2001.

[5] C. Shen, H. Li, and M.J. Brooks. Supervised dimensionality reduction via sequential semidefinite programming. *PR*, 41(12):3644–3652, 2008.

[6] Shuicheng Yan, Dong Xu, Benyu Zhang, Hong-Jiang Zhang, Qiang Yang, and Stephen Lin. Graph embedding and extensions: a general framework for dimensionality reduction. *PAMI*, 29(1):40–51, 2007.

[7] M. Zhu and A.M. Martinez. Subclass discriminant analysis. *PAMI*, 28 (8):1274–1286, 2006.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.