Heteroscedastic Max-min Distance Analysis

Bing Su\textsuperscript{1}, Xiaoging Ding\textsuperscript{1}, Changsong Liu\textsuperscript{1}, Ying Wu\textsuperscript{2}

\textsuperscript{1}Department of Electronic Engineering, Tsinghua University. \textsuperscript{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 between classes, which will cause the so-called “class separation” problem.

Max-min distance analysis 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 degrades the performance of MMDA as shown in Fig. 1(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 scatterers. Whitened HMMDA (WHMMDA) maximizes the minimal pairwise Chernoff 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:

$$d_{Cij} = (\hat{\mu}_i - \hat{\mu}_j)^T \hat{S}^{-1}_{Cij} (\hat{\mu}_i - \hat{\mu}_j) + \frac{1}{\alpha_i (1 - \alpha_j)} \log \left( \frac{S_{ij}}{\hat{S}_{ij}} \right)$$

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

$$S_{Cij} = \hat{S}^{-1}_{ij} (\hat{\mu}_i - \hat{\mu}_j)(\hat{\mu}_i - \hat{\mu}_j)^T \hat{S}^{-1}_{ij} + \frac{1}{\alpha_i (1 - \alpha_j)} (\log \hat{S}_{ij} - \alpha_i \log \hat{S}_i - (1 - \alpha_i) \log \hat{S}_j)$$

Hence the objective of WHMMDA is:

$$\max_{W_2} \min_{1 \leq i < j \leq C} \left( (p_i p_j)^{-1} tr(W_2^T S_{Cij} W_2) \right)$$

subject to $W_2^T W_2 = I_p$

(3) can be optimized by introducing variable $Z = W W^T$ and relaxing the feasible set of $Z$ 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:

$$\max_{W_2} \min_{1 \leq i < j \leq C} \left( (p_i p_j)^{-1} tr(W_2^T S_{Cij} W_2) \right)$$

subject to $W_2^T W_2 = I_p$

(4) can be solved by the bisection search strategy proposed in [5] by relaxing the feasible set of $Z$ and deriving the upper and lower bound of the auxiliary value $\alpha$.

Different formulations of $S_{Cij}$ and $S_{ij}$ 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 scatterers can be used to force adjacent samples to be close:

$$\Sigma_i = \sum_{x_k \in \text{class } i} \sum_{x_q \in N_i^x (x_k)} (x_k - x_q)(x_k - x_q)^T$$

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

$$S_{Cij}^m = (\Sigma_{ij}^m)^{-1/2} (m_i - m_j)(m_i - m_j)^T (\Sigma_{ij}^m)^{-1/2} + \frac{1}{\alpha_i (1 - \alpha_j)} (\log \Sigma_{ij}^m - \alpha_i \log \Sigma_i^m - (1 - \alpha_j) \log \Sigma_j^m)$$

(6) can be obtained by introducing variable $Z = W W^T$ and relaxing the feasible set of $Z$ to its convex hull following [1].