LOCALITY SENSITIVE DISCRIMINANT ANALYSIS

Authors: D. Cai, X. He, K. Zhou, J. Han, H. Bao

Reporter: Yunfei Wang

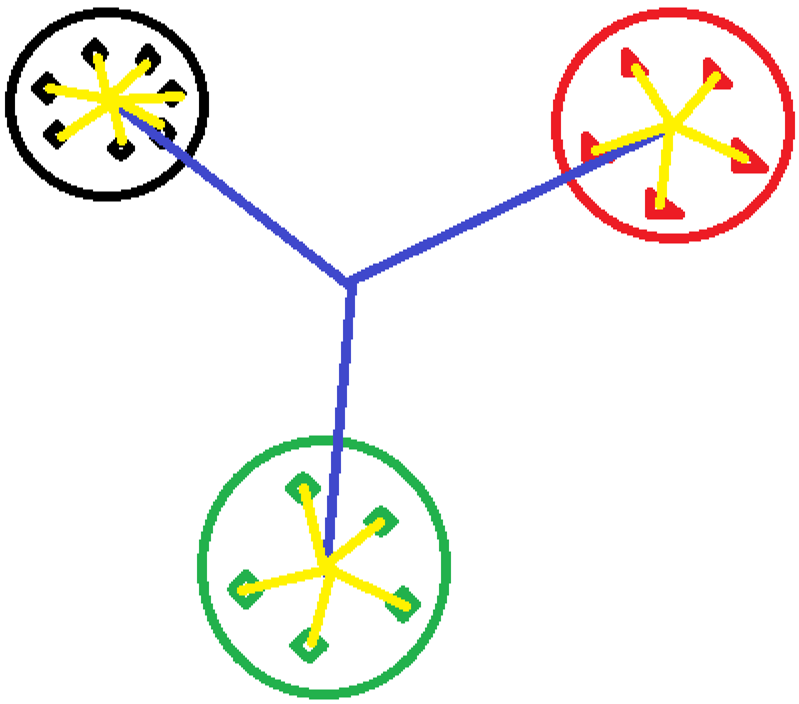
Sep. 28,2012

Outlines

- Linear Discriminant Analysis(LDA)
- Locality Sensitive Discriminant Analysis(LSDA)
- Marginal Fisher Analysis(MFA)
- Experimental Results
- Summary-Comparison between LDA,LSDA and MFA

Linear Discriminant Analysis

- Maximizing the ratio of inter-class variance to the intra-class variance in any particular data set thereby guaranteeing maximal separability.



LDA-Objective Function

$$\mathbf{a}_{opt} = \arg\max_{\mathbf{a}} \frac{\mathbf{a}^T S_b \mathbf{a}}{\mathbf{a}^T S_w \mathbf{a}}$$

$$S_b = \sum_{i=1}^{c} m_i (\boldsymbol{\mu}^i - \boldsymbol{\mu})(\boldsymbol{\mu}^i - \boldsymbol{\mu})^T$$

$$S_w = \sum_{i=1}^{c} \left(\sum_{j=1}^{m_i} (\mathbf{x}_j^i - \boldsymbol{\mu}^i)(\mathbf{x}_j^i - \boldsymbol{\mu}^i)^T \right)$$

$$S_b \mathbf{a} = \lambda S_w \mathbf{a}$$

Drawbacks of LDA

- It's motivated by the assumption than the data of each class is Gaussian distributed, which is not always satisfied in real world problems. Without Gaussian distribution assumption, the inter-class scatter can't well characterize the separability of the data of different class.
- LDA fails to discover the intrinsic local geometrical structure of the data manifold while trying to preserve the global class relationship.

LSDA

- Estimating geometrical and discriminant properties of the submanifold form random points lying on this unknown submanifold.

- Finding a projection which maximizes the margin between data points from different class at each local area.

LSDA-Neighbor Graph

- Global neighbor graph

$$W_{ij} = \begin{cases} 1, & \text{if } \mathbf{x}_i \in N(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in N(\mathbf{x}_i) \\ 0, & \text{otherwise.} \end{cases}$$

- Inter-class neighbor graph

$$W_{b,ij} = \begin{cases} 1, & \text{if } \mathbf{x}_i \in N_b(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in N_b(\mathbf{x}_i) \\ 0, & \text{otherwise.} \end{cases}$$

- Intra-class neighbor graph

$$W_{w,ij} = \begin{cases} 1, & \text{if } \mathbf{x}_i \in N_w(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in N_w(\mathbf{x}_i) \\ 0, & \text{otherwise.} \end{cases}$$

LSDA-Objective Functions

- After mapping the intra-class graph and inter-class graph to a projection vector ***a***, we get a map $\boldsymbol{y} = \left(y_1, y_2, \ldots y_m\right)^T$

- Minimizing the intra-class scatter:

$$\min \sum_{ij} (y_i - y_j)^2 W_{w,ij}$$

- Maximizing the inter-class scatter:

$$\max \sum_{ij} (y_i - y_j)^2 W_{b,ij}$$

LSDA-Optimal Linear Embedding

$$\begin{aligned}
&\frac{1}{2}\sum_{ij}(y_i - y_j)^2 W_{w,ij} \\
=\ &\frac{1}{2}\sum_{ij}\left(\mathbf{a}^T\mathbf{x}_i - \mathbf{a}^T\mathbf{x}_j\right)^2 W_{w,ij} \\
=\ &\sum_{i}\mathbf{a}^T\mathbf{x}_i D_{w,ii}\mathbf{x}_i^T\mathbf{a} - \sum_{ij}\mathbf{a}^T\mathbf{x}_i W_{w,ij}\mathbf{x}_j^T\mathbf{a} \\
=\ &\mathbf{a}^T X D_w X^T\mathbf{a} - \mathbf{a}^T X W_w X^T\mathbf{a}
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{2}\sum_{ij}(y_i - y_j)^2 W_{b,ij} \\
=\ &\frac{1}{2}\sum_{ij}\left(\mathbf{a}^T\mathbf{x}_i - \mathbf{a}^T\mathbf{x}_j\right)^2 W_{b,ij} \\
=\ &\mathbf{a}^T X(D_b - W_b)X^T\mathbf{a} \\
=\ &\mathbf{a}^T X L_b X^T\mathbf{a}
\end{aligned}$$

- Constraint: $\mathbf{y}^T D_w \mathbf{y} = 1 \Rightarrow \mathbf{a}^T X D_w X^T \mathbf{a} = 1$

- Final Optimization Problem:

$$\arg\max_{\substack{\mathbf{a} \\ \mathbf{a}^T X D_w X^T\mathbf{a} = 1}} \mathbf{a}^T X\big(\alpha L_b + (1-\alpha)W_w\big)X^T\mathbf{a}$$

Marginal Fisher Analysis

- MFA has the same goal with LDA, but uses different criterion to characters the intra-class compactness and the inter-class separability.
  - Intra-class compactness: sum of distances between each point and its nearest neighbors of the same class.
  - Inter-class separability: sum of distances between margin points and their neighboring points of different classes.

- MFA overcomes the limitation of Gaussian distribution assumption of LDA.

MFA-Criterions

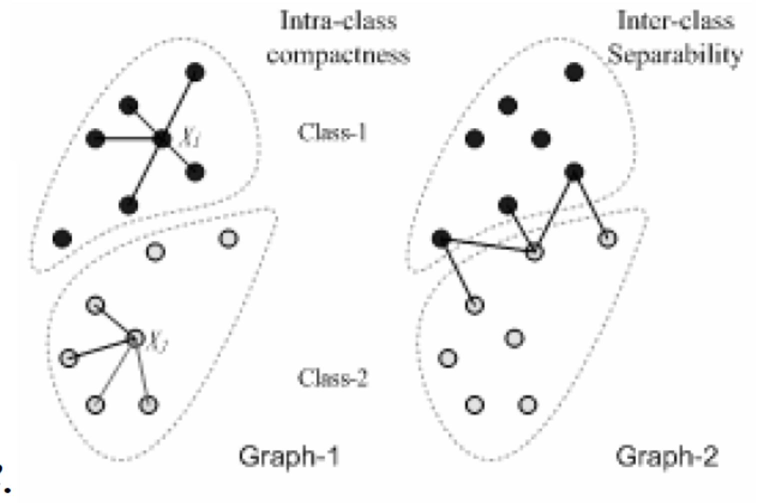


Figure 3: The two adjacency graphs for Marginal Fisher Analysis. Note that the left adjacency graph only plots the connection edges for one sample in each class for ease of understanding.

- Intra-class compactness:

$$\tilde{S}_c = \sum_i \sum_{i \in N_{k_1}^+(j) \, or \, j \in N_{k_1}^+(i)} \| w^{\mathrm{T}} x_i - w^{\mathrm{T}} x_j \|^2$$

$$= 2 w^{\mathrm{T}} X (D^c - W^c) X^{\mathrm{T}} w$$

$$W_{ij}^c = 1 \quad if \ j \in N_{k_1}^+(i) \ or \ i \in N_{k_1}^+(j); \ 0, else.$$

- Inter-class separability:

$$\tilde{S}_m = \sum_i \sum_{(i,j) \in P_{k_2}(l_i) \, or \, (j,i) \in P_{k_2}(l_j)} \| w^{\mathrm{T}} x_i - w^{\mathrm{T}} x_j \|^2$$

$$= 2 w^{\mathrm{T}} X (D^m - W^m) X^{\mathrm{T}} w$$

$$W_{ij}^m = 1 \ \ if \ (i, j) \in P_{k_2}(l_i) \ or \ (j, i) \in P_{k_2}(l_j); \ \ 0, else.$$

MFA-Algorithmic Procedure

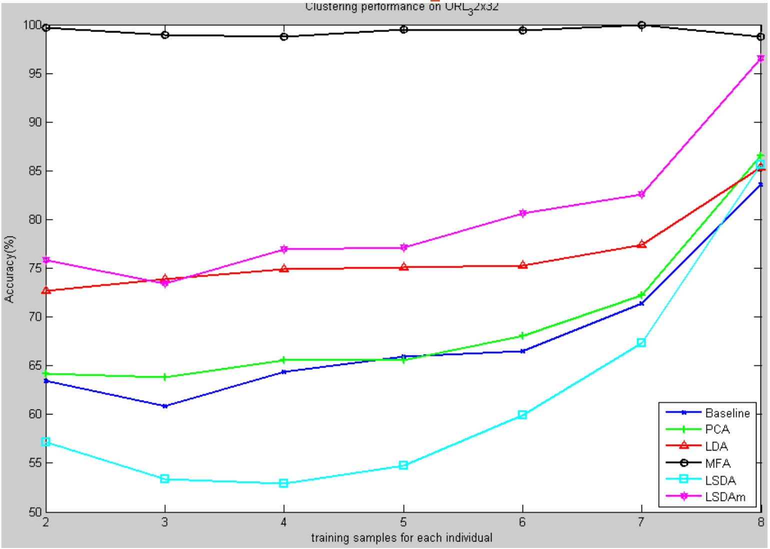
- PCA projection: projecting the data into PCA subspace by remaining N-N_c dimensions. W_{PCA} denotes the transformation matrix of PCA.
- Constructing the intra-class compactness and inter-class separability graphs W^c and W^m.
- Marginal Fisher Criterion:

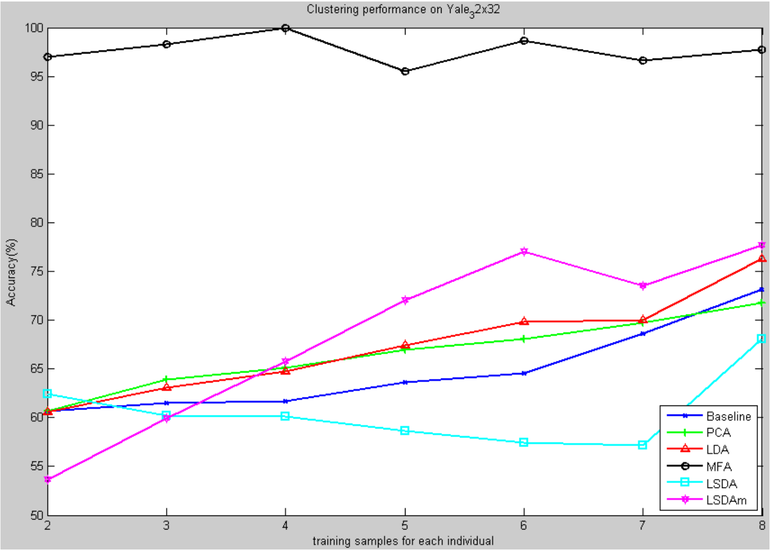
$$w^* = \arg\min_w \frac{w^T X(D^c - W^c)X^T w}{w^T X(D^m - W^m)X^T w}$$

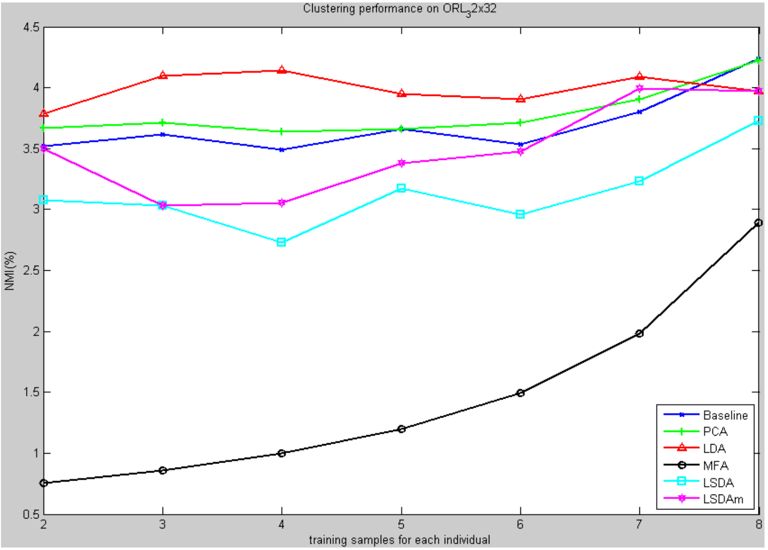
- Output the final linear projection direction:

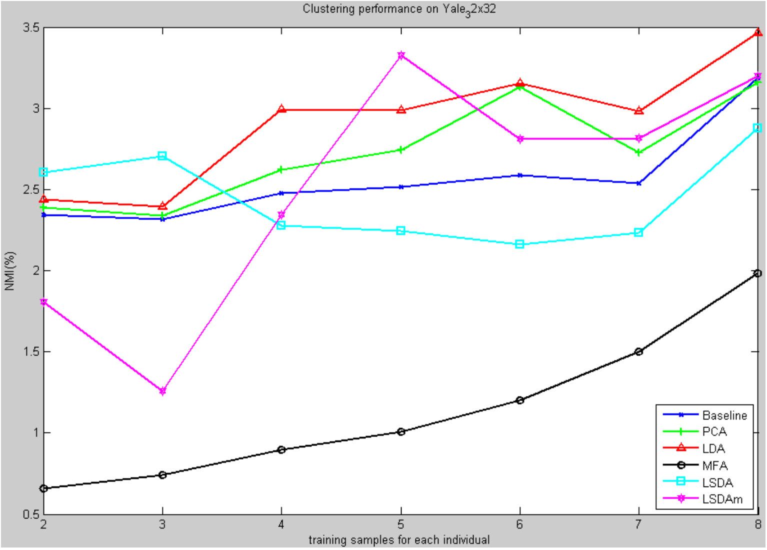
$$w = W_{PCA} w^*$$

Experimental Results









Comparison

- Sharing points: compact the points of each class and separate the different classes as much as possible.

- LDA: mainly focus on the global geometry structure.

- LSDA: takes the local submanifold into consideration.

- MFA: takes advantage of a new criterion to estimating the intra-class and inter-class scatter.