A Family of Maximum Margin Criterion for Adaptive Learning

Miao Cheng

E-mail: mewcheng@gmail.com

ICONIP, 2018 DEC

Outline

Outline

Background

Subspace Learning Methods

- Powerful for pattern analysis
- Benchmark for further development
- Linear Classifiers

Background

Subspace Learning Methods

- Principle Component Analysis (PCA)
- Linear Discriminant Analysis (LDA)
- Maximum Margin Criterion (MMC)

Background

General Problems

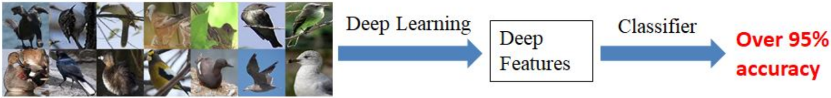
- Large sizes of data
- Efficient calculation

Solutions

- Pre-step
- Direct approaches

Background

Bird200 Dataset



Deep learning used as a feature extraction method

Outline

MMC

Use substraction objective

$$J_{MMC} = w^T \left(S_b - \gamma S_w \right) w \tag{1}$$

Singularity problem of S_w (denominator in LDA) can be avoided

MMC

Use substraction objective

$$J_{MMC} = w^T \left(S_b - \gamma S_w \right) w \tag{2}$$

Singularity problem of S_w (denominator in LDA) can be avoided

Limitation

Efficient calculation is suspended

MMC

Comments

"Supposed that there is a PCA pre-step, and whole data energy is preserved. Then, there would be no lost discriminant power."

Outline

Direct MMC

The learned subspace is spanned by samples $X \in R^{d \times n}$

$$\begin{aligned} J &= w^T X^T \left(S_b - \gamma S_w\right) X w \\ &= w^T K (L_b - \gamma L_w) K w \end{aligned} \tag{3}$$

$$X E \Lambda^{-\frac{1}{2}} = U S V^T \tag{4}$$

Direct MMC

The learned subspace is spanned by samples $X \in R^{d \times n}$

$$\begin{aligned} J &= w^T X^T \left(S_b - \lambda S_w\right) X w \\ &= w^T K(L_b - \gamma L_w) K w \end{aligned} \tag{5}$$

$$X E \Lambda^{-\frac{1}{2}} = U S V^T \tag{6}$$

Complexity

- The calculation complexity mainly depends on n, $O\left(n^3\right)$
- The n can be very large

Outline

Random MMC

Randomly select t samples $A = [x_{a1}, x_{a2}, \cdots, x_{at}]$

$$J = M\left(L_b - \gamma L_w\right)M^T, \quad M = A^T X \tag{7}$$

Random MMC

Randomly select t samples $A = [x_{a1}, x_{a2}, \cdots, x_{at}]$

$$J = M\left(L_b - \gamma L_w\right)M^T, \quad M = A^T X \tag{8}$$

Complexity

The calculation complexity mainly depends on t

Outline

Layered MMC

Inspired by layered structures in extreme learning machine (ELM)

$$x \mapsto h(x), \quad R^d \mapsto R^g \tag{9}$$

Layered MMC

$$\begin{aligned} x &\mapsto h(x), \quad R^d \mapsto R^g \\ h(x) &= Bx \end{aligned} \tag{10}$$

$$J = M(L_b - \lambda L_w)M^T, \quad M = A^T BX, \quad B = PP^T \tag{11}$$

$2D^2$MMC

Efficient handling of 2D data

$$y = P^T x Q, \quad x \in R^{d_1 \times d_2} \tag{12}$$

$$\begin{aligned}
S_{bl} &= \sum_{i=1}^{c} n_i \left(m_i^2 - m^2\right) \left(m_i^2 - m^2\right)^T \\
S_{wl} &= \sum_{i=1}^{c} \sum_{j=1}^{n_i} \left(x_{ij}^2 - m_i^2\right) \left(x_{ij}^2 - m_i^2\right)^T \\
S_{br} &= \sum_{i=1}^{c} n_i \left(m_i^2 - m^2\right)^T \left(m_i^2 - m^2\right) \\
S_{wr} &= \sum_{i=1}^{c} \sum_{j=1}^{n_i} \left(x_{ij}^2 - m_i^2\right)^T \left(x_{ij}^2 - m_i^2\right)
\end{aligned} \tag{13}$$

$2D^2$MMC

Efficient handling of 2D data

$$y = P^T x Q, \quad x \in R^{d_1 \times d_2} \tag{14}$$

$$\begin{aligned} J(p) &= p^T (S_{bl} - \gamma S_{wl})\, p \\ J(q) &= q^T (S_{br} - \gamma S_{wr})\, q. \end{aligned} \tag{15}$$

MMC Network

- MMC development of LDA network proposed in PCANet (T. H. Chan, K. Jia)
- Pixel patches of each image

Example

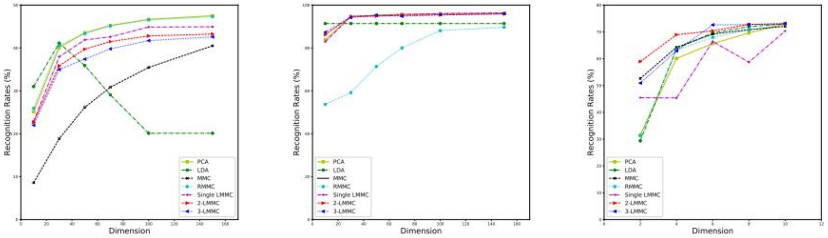
LBP features

Outline

Data sets

- SUN: 512 d VGG-16 features, 100 classes, half for training and testing
- MNIST: Simple sparse coding features, 10 classes, 2000 vs. 500
- STL-10: deep features of target coding, 2000 vs. 500

Experimental results



(a) (b) (c)

Fig. 1. The results of different methods on three data sets. (a) Experimental results on SUN database. (b) Experimental results on MNIST database. (c) Experimental results on STL-10 data set.

Data sets

- ALOI: 50 categories, 18 vs. 54 images in each class
- MNIST: 10 classes, 2000 vs. 500 for each class

Experimental results

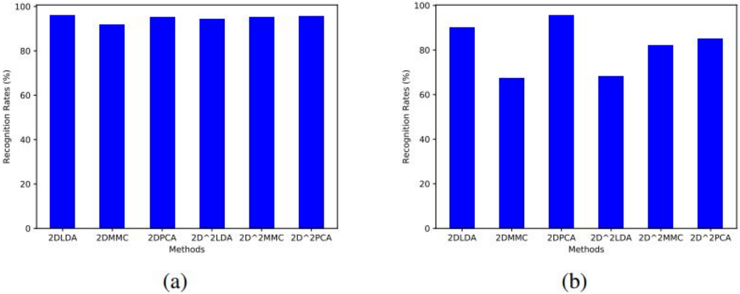


Fig. 2. The results of different 2D methods on three data sets. (a) Experimental results on ALOI database. (b) Experimental results on MNIST database.

Data sets

- ALOI: 30 categories, size of 30*30
- MNIST: 10 classes, 100 vs. 100 for each class

Experimental results

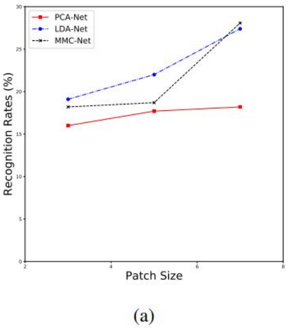


Fig. 3. The results of different 2D methods on three data sets. (a) Experimental results on MNIST data set. (b) Experimental results on ALOI data set.

Thank you