CH3.8 Component Analysis and Discriminants

Yuxin Ding of HIT Shenzhen

Department of Computer science
HIT Shenzhen

Fall 2006

CONTENTS

Component Analysis and Discriminants

- Cope with the problem of excessive dimensionality: reduce the dimensionality by combining features.
- approaches
  - Principal Component Analysis (PCA)
  - Fisher Linear Discriminant
  - Multiple Discriminant Analysis

Principal Component Analysis

- Seek a projection that best represents the data in a least-squares sense.
- Three cases of projecting the data set with the high dimension onto a lower dimension:

  Zero-dimensional representation

  One-dimensional representation

  d-dimensional representation

Zero-Dimensional Representation

Representing all of the vectors in a data set $D = \{x_1, x_2, \cdots, x_n\}$ by a single vector x_0. Define the squared-error criterion function $J_0(x_0)$:

$$J_0(x_0) = \sum_{k=1}^{n} (\|x_0 - x_k\|)^2$$

GOAL: Seek the value x_0 that minimizes J_0.

Zero-Dimensional Representation(cond.) I

The sample mean of set $D = \{x_1, x_2, \cdots, x_n\}$:

$$m = \frac{1}{n} \sum_{k=1}^{k} x_k$$

Zero-Dimensional Representation(cond.) II

then we obtain:

$$
\begin{aligned}
J_0(x_0) &= \sum_{k=1}^{n} \|(x_0 - m) - (x_k - m)\|^2 \\
&= \sum_{k=1}^{n} \|(x_0 - m)\|^2 - 2\sum_{k=1}^{n}(x_0 - m)^t(x_k - m) + \sum_{k=1}^{n} \|(x_k - m)\|^2 \\
&= \sum_{k=1}^{n} \|(x_0 - m)\|^2 - 2\sum_{k=1}^{n}(x_0 - m)^t(x_k - m) + \sum_{k=1}^{n} \|(x_k - m)\|^2 \\
&= \sum_{k=1}^{n} \|(x_0 - m)\|^2 + \sum_{k=1}^{n} \|(x_k - m)\|^2
\end{aligned}
$$

J_0 is minimized by the choice $x_0 = m$

One-Dimensional Representation

Projecting all of the vectors in a data set $D = \{x_1, x_2, \cdots, x_n\}$ onto a line running through the sample space.

The equation of the line:

$$x = m + \alpha e$$

Where e be a unit vector.
We represent x_k by $m + \alpha_k e$

One-Dimensional Representation (cond.) I

Find an optimal set of coefficients ak by minimizing the squared-error criterion function:

$$
\begin{aligned}
J_1(a_1, \cdots, a_n, e) &= \sum_{k=1}^{n} \|(m + a_k e) - x_k\|^2 \\
&= \sum_{k=1}^{n} \|a_k e - (x_k - m)\|^2 \\
&= \sum_{k=1}^{n} a_k^2 \|e\|^2 - 2 \sum_{k=1}^{n} a_k e^t (x_k - m) + \sum_{k=1}^{n} \|(x_k - m)\|^2
\end{aligned} \tag{1}
$$

One-Dimensional Representation (cond.) II

Recognizing $\|e\| = 1$ and Partially differentiating w.r.t a_k :

$$a_k = e^t(x_k - m)$$

So far, we only obtain the coefficient a_k for the vector x_k projected onto the line in the direction of e.

Next step is to determine the line direction e.

Define scatter-matrix S:

$$S = \sum_{k=1}^{n}(x_k - m)(x_k - m)^t$$

One-Dimensional Representation (cond.) III

From following two Eqs.

$$J_1(a_1, \cdots, a_n, e) = \sum_{k=1}^{n} a_k^2 - 2\sum_{k=1}^{n} a_k e^t (x_k - m) + \sum_{k=1}^{n} \|(x_k - m)\|^2$$

$$a_k = e^t (x_k - m)$$

One-Dimensional Representation (cond.) IV

We obtain:

$$
\begin{aligned}
J_1(e) &= \sum_{k=1}^{n} a_k^2 \|e\|^2 - 2\sum_{k=1}^{n} a_k^2 + \sum_{k=1}^{n} \|(x_k - m)\|^2 \\
&= -\sum_{k=1}^{n} [e^t(x_k - m)]^2 + \sum_{k=1}^{n} \|x_k - m\|^2 \\
&= -\sum_{k=1}^{n} e^t(x_k - m)(x_k - m)^t e + \sum_{k=1}^{n} \|x_k - m\|^2 \\
&= -e^t S e + \sum_{k=1}^{n} \|x_k - m\|^2
\end{aligned}
$$

One-Dimensional Representation (cond.) V

Use the Lagrange multiplier method to maximize $e^t Se$ subject to the constraint $\|e\|^2 = 1$. Letting λ be the undetermined multiplier. We differentiate $u = e^t Se - \lambda e^t e + \lambda$

w. r. t. e to obtain:

$$\frac{\partial u}{\partial e} = 2Se - 2\lambda e$$

e must be the eigenvector corresponding to the largest eigenvalue of the scatter matrix S.

$$Se = \lambda e$$

d'-Dimensional Representation

Projecting all of the vectors in a data set $D = \{x_1, x_2, \cdots, x_n\}$, onto d dimensional space.

$$x = m + \sum_{i=1}^{d'} a_i e_i$$

Where $d' <= d$, we can obtain that the criterion function:

$$J_{d'} = \sum_{k=1}^{n} \|(m + \sum_{i=1}^{d'} a_{k_i} e_i) - x_k\|^2$$

is minimized when the vectors $e1, e2$, are the eigenvectors of the scatter matrix having the largest eigenvalues.

Fisher Linear Discriminant I

- PCA: seeks directions that are efficient for presentation.
- Discriminant analysis: Seeks directions that are efficient for discrimination.

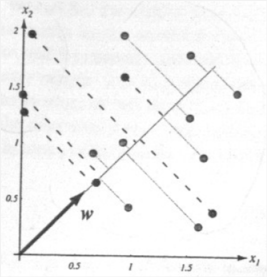
  Suppose that we have a set of n d-dimensional samples $x_1, x_2, \cdots, x_n$ n_1 samples in the subset $D1$ labeled Ω_1 and n_2 samples in the subset D_2 labeled Ω_2 we form a linear combination of the components of x as:

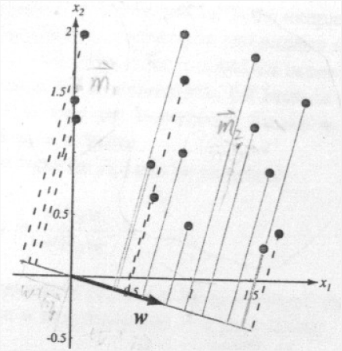
$$y = w^t x$$

Fisher Linear Discriminant II

A corresponding set of $\{y_1, y_2, \cdots, y_n\}$ divided into the subset $Y1$ and $Y2$. if $\|w\| = 1$, each y_i is the projection of the corresponding onto a line in the direction w.

Fisher Linear Discriminant III





The figure on the right shows greater separation between subsets, one set of the points with dashed line, another with solid line.

Fisher Linear Discriminant (cond.) I

- Find the best direction w that we will obtain accurate classification.

  A measure of the separation between the projected points is the difference of the sample means.

  If m_i is the d-dimensional sample mean from D_i given by $m_i = \frac{1}{n_i}\sum_{X \in D_i} X$ the sample mean from the projected points Y_i given by:

$$\tilde{m}_i = \frac{1}{n_i}\sum_{y \in Y_i} y = \frac{1}{n_i}\sum_{x \in D_i} w^t x = w^t m_i$$

Fisher Linear Discriminant (cond.) II

the difference of the projected sample means is:

$$\|\tilde{m}_1 - \tilde{m}_2\| = \|w^t(m_1 - m_2)\|$$

Goal: maximize this difference.

Define scatter for projected samples labeled ω_i:

$$\tilde{s}_i^2 = \sum_{y \in Y_i} (y - \tilde{m}_i)^2$$

Fisher Linear Discriminant (cond.) III

The FLD employs that linear function $w_t x$ for which the following criterion function is maximun:

$$J(W) = \frac{\|\tilde{m}_1 - \tilde{m}_2\|^2}{\tilde{s}_1{}^2 + \tilde{s}_2{}^2}$$

$\tilde{s}_1{}^2 + \tilde{s}_2{}^2$ is called *the total within-class scatter*.

To obtain $J(w)$ as explicit function of w, we define scatter matrices $S_i(i = 1, 2)$ and S_w by:

$$S_i = \sum_{x \in D_i} (x - m_i)(x - m_i)^t \qquad S_W = S_1 + S_2$$

Fisher Linear Discriminant (cond.) IV

From the following Eqs.

$$y = w^t x \tag{2}$$

$$\tilde{m}_i = \frac{1}{n_i} \sum_{y \in Y_i} y = \frac{1}{n_i} \sum_{x \in D_i} w^t x = w^t m_i \tag{3}$$

$$\tilde{s}_i^2 = \sum_{y \in Y_i} (y - \tilde{m}_i)^2 \tag{4}$$

Fisher Linear Discriminant (cond.) V

We can write

$$\tilde{s}_i^2 = \sum_{x \in D_i} (w^t x - w^t m_i)^2$$
$$= \sum_{x \in D_i} w^t (x - m_i)(x - m_i^t w)$$
$$= w^t S_i w$$

We obtain:

$$\tilde{s_1^2} + \tilde{s_2^2} = w^t S_W w$$

From the Eq. $\|\tilde{m}_1 - \tilde{m}_2\| = \|w^t(m_1 - m_2)\|$

Fisher Linear Discriminant (cond.) VI

We obtain:

$$\begin{aligned}\|\tilde{m}_1 - \tilde{m}_2\| &= (w^t m_1 - w^t m_2)^2 \\ &= w^t (m_1 - m_2)(m_1 - m_2)^t w \\ &= w^t S_B w\end{aligned}$$

where: $S_B = (m_1 - m_2)(m_1 - m_2)^t$

In terms of S_B and S_w, $J(w)$ can be written as:

$$J(w) = \frac{w^t S_B w}{w^t S_W w}$$

Fisher Linear Discriminant (cond.) VII

A vector w that maximizes $J(w)$ must satisfy:

$$S_B w = \lambda S_W w$$

In the case that S_w is nonsingular.

$$S_W^{-1} S_B w = \lambda w$$

Due to the fact that $S_B w$ is always in the direction of $m_1 - m_2$.

we obtain:

$$w = S_W^{-1}(m_1 - m_2)$$

Multiple Discriminant Analysis I

- For the c-class problem, the nature generalization of Fisher's linear discriminant involves $c - 1$ discriminant functions
- The projection is from a d-dimensional space to a $(c - 1)$-dimensional space.
- It is tacitly assumed that $d >= c$

Multiple Discriminant Analysis II

The generalization for the within-class scatter matrix is :

$$S_W = \sum_{i=1}^{c} S_i$$

where, as before:

$$S_i = \sum_{x \in D_i} (x - m_i)(x - m_i)^t$$

and :

$$m_i = \frac{1}{n_i} \sum_{x \in D_i} x$$

Multiple Discriminant Analysis III

The generalization for S_B is not so obvious.

- Define the total mean vector m and total scatter matrix S_T as follows:

$$m = \frac{1}{n}\sum_{x} x = \frac{1}{n}\sum_{i=1}^{c} n_i m_i$$

$$S_T = \sum_{x}(x-m)(x-m)^t$$

Multiple Discriminant Analysis IV

Then it follows that:

$$
\begin{aligned}
S_T &= \sum_{i=1}^{c} \sum_{x \in D_i} (x - m_i + m_i - m)(x - m_i + m_i - m)^t \\
&= \sum_{i=1}^{c} \sum_{x \in D_i} (x - m_i)(x - m_i)^t + \sum_{i=1}^{c} \sum_{x \in D_i} (m_i - m)(m_i - m)^t
\end{aligned}
$$

- Define the between-class scatter matrix as:

$$S_B = \sum_{i=1}^{c} n_i (m_i - m)(m_i - m)^t$$

Multiple Discriminant Analysis V

- The total scatter is: $S_T = S_W + S_B$
- The projection from d-dimensional space to $(c-1)$-dimensional space is accomplished by $(c-1)$ discriminant functions.

$$y_i = w_i^t x (i = 1, \cdots, c-1)$$

Multiple Discriminant Analysis VI

- Define vector Y from y_i, d-by-$(c-1)$ matrix W from vector w_i, the projection can be written as :

$$Y = W^t x$$

Multiple Discriminant Analysis VII

- The samples $x_1, \cdots, x_n$ project to a corresponding set $y_1, \cdots, y_n$,which can be described by their own mean vectors and scatter matrix. Define:

$$\tilde{m}_i = \frac{1}{n_i} \sum_{y \in Y_i} y$$

$$\tilde{m} = \frac{1}{n} \sum_{i=1}^{c} n_i \tilde{m}_i$$

$$\tilde{S_W} = \sum_{i=1}^{c} \sum_{y \in Y_i} (y - \tilde{m}_i)(y - \tilde{m}_i)^t$$

$$\tilde{S_B} = \sum_{i=1}^{c} n_i (\tilde{m}_i - \tilde{m})(\tilde{m}_i - \tilde{m})^t$$

Multiple Discriminant Analysis VIII

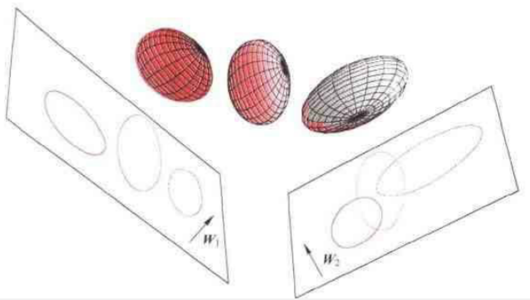
it is a straightforward matter to show that:

$$\tilde{S_W} = W^t S_W W$$

$$\tilde{S_B} = W^t S_B W$$

Multiple Discriminant Analysis IX

These equations shows how the within-class and between-class scatter matrices transformed by the projection.



Three-dimensional distributions are projected onto two-dimensional subspaces, described by vectors w_1 and w_2

Multiple Discriminant Analysis X

GOAL: Find transformation matrix W maximizes the ratio :

$$J_W = \frac{\tilde{S_B}}{\tilde{S_W}} = \frac{W^t S_B W}{W^t S_W W}$$

Find W:

- The columns of an optimal W are the generalized eigenvectors that correspond to the lagest eigenvalues in:

$$S_B w_i = \lambda_i S_W w_i$$

Multiple Discriminant Analysis XI

- We can find the eigenvalues as the root of the characteristic polynomial :

$$|S_B - \lambda_i S_W| = 0$$

and then solve:

$$(S_B - \lambda_i S_W) w_i = 0$$

to find the eigenvectors w_i

THE END