EE 7403 LEC 12 Feature Extraction / Dimension Reduction v. Machine Learning

1. Intro

Pattern Recognition is a series of processes that reduce the dimensionality and the variation of samples for the same class and keep their discrimination for different classes.

$\Rightarrow$ Dimensionality Reduction & Discriminative info extraction.

feature extraction & dimensionality reduction
- human expert knowledge: fingerprint (minutia points)
- Image local structure: corners, blobs, interesting points
- Image global structure
- Machine Learning from training database.

2. Feature Extraction Based on Image Global Structure

Transform image $f(x,y)$ to feature $g(u,v)$

$$g(u,v) = T\{f(x,y)\}$$
$$= \iint w(u,v,x,y) f(x,y)\,dx\,dy \quad \text{linear transform}$$
$$\Rightarrow \sum_1^h \sum_1^w w(u,v,x,y) f(x,y) \quad \text{for digital image}$$
$$\Rightarrow \sum_1^h \sum_1^w e^{-2\pi j(ux+vy)} f(x,y) \quad \text{Fourier transform}$$
$$\Rightarrow \sum_1^h \sum_1^w x^u y^v f(x,y) \quad \text{moments computing}$$
$$\Rightarrow f = W^T x \quad \text{vector-matrix representation}$$

Polar Complex Exponential Transform (PCET)

$$g(u,v) = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 e^{-j(2\pi u r^2 + v\theta)} f(r,\theta)\,dr\,d\theta$$

3. Principal Component Analysis unsupervised

Given q n-dim training samples:

$$x_1, x_2, \dots, x_q$$

$$\mu = \frac{1}{q} \sum_{i=1}^{q} x_i$$

$\tilde{x}_i = x_i - \mu \qquad X = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_q]$

use 1-dim to represent $\tilde{x}_i$ best

$a_i = \phi^T \tilde{x}_i \quad (\phi: \|\phi\|^2 = \phi^T\phi = 1)$

reduce computational complexity

lost info maybe x important for data representation but ↓ for discriminating possibly.

EE7403 LEC 12

the best ϕ makes reconstruction error minimum

$$\varepsilon^2 \sum_{i=1}^{q} \| \tilde{x}_i - a_i\phi \|^2 \Rightarrow \text{minimum}$$

$$\Rightarrow \varepsilon^2 = \sum \| \tilde{x}_i - a_i\phi \|^2 = \sum (\tilde{x}_i - a_i\phi)^T(\tilde{x}_i - a_i\phi) = \sum \left(\tilde{x}_i^T\tilde{x}_i - 2a_i\phi^T\tilde{x}_i + a_i^2\phi^T\phi\right)$$

$$= \sum (\tilde{x}_i^T\tilde{x}_i - 2a_i^2 + a_i^2) = \sum (\tilde{x}_i^T\tilde{x}_i - a_i^2)$$

$$= \sum \|\tilde{x}_i\|^2 - \sum a_i^2 = \sum \|\tilde{x}_i\|^2 - \sum (\phi^T\tilde{x}_i)(\phi^T\tilde{x}_i)^T$$

$$= \sum \|\tilde{x}_i\|^2 - \phi^T\left(\sum \tilde{x}_i\tilde{x}_i^T\right)\phi$$

the sample covariance matrix of all training data

$$\boxed{S^t} = \frac{1}{q}\sum_{i=1}^{q}(x_i - \mu)(x_i - \mu)^T = \underline{\frac{1}{q}\sum_{i=1}^{q}\tilde{x}_i\tilde{x}_i^T} \rightarrow \frac{1}{q}XX^T$$

S^t is total scatter matrix

the rank of S^t is $\min(q-1, n)$ at most

to minimize $\varepsilon^2 = \sum\|\tilde{x}_i\|^2 - q\phi^T S^t \phi$

use Lagrange optimization

S^t is symmetric

$S^t = (S^t)^T$

$$f(\phi, \lambda) = \phi^T S^t \phi - \lambda(\phi^T\phi - 1) \Rightarrow \text{maximum}$$

$$\frac{\partial f}{\partial \phi} = 2S^t\phi - 2\lambda\phi = 0$$

$$S^t\phi = \lambda\phi$$

$\therefore$ the solution is the the $\boxed{\text{eigenvalues}}$ and $\boxed{\text{eigen-vectors}}$ of S^t. (and eigenvectors ϕ satisfies $\phi^T\phi$)

notes:

require $f(x,y)$ reaches maximum when $g(x,y) = C$

$$L(x, y, \lambda) = f(x,y) + \lambda \cdot (g(x,y) - C)$$

whose pole values contains pole values of $f(x,y)$

$$\Rightarrow \nabla L = 0$$

$$\varepsilon^2 = \sum\|x_i\|^2 - q\phi^T S^t \phi = \sum\|x_i\|^2 - q\phi^T\lambda\phi = \sum\|x_i\|^2 - \lambda q$$

Reducing the x_i into lower-dim m-dim y_i by

$$\boxed{y_i = \phi^T(x_i - \mu)} \quad \phi = [\underbrace{\phi_1, \phi_2, \ldots, \phi_m}_{m \text{ largest eigenvalues of } S^t}] \quad m < n$$

reconstruct

$$\boxed{\hat{x}_i = \phi y_i + \mu}$$

the reconstruction error $E[\|\Delta\|^2] = E[\Delta^T\Delta] = \sum_{k=m+1}^{n} \lambda_k$ (λ_k are λ storted in descending order)

EE7403 LEC 12

4. Eigen decomposition

eigenvalue and eigenvector :

$$\Sigma \phi_i = \lambda_i \phi_i \quad i = 1, 2, \ldots, n$$

if Σ is $\boxed{\text{symmetric}}$ eigenvectors corresponding to the distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ are orthogonal

Take the unit length of $\phi_1, \phi_2, \ldots, \phi_n$

$$\phi_i^T \phi_j = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

(orthogonal + unit length = orthonormal)

let ϕ be the orthonormal matrix formed by eigenvectors

$$\phi = [\phi_1\ \phi_2, \ldots, \phi_n]$$

Obviously $\phi^T \phi = I$ $\quad \therefore \phi^{-1} = \phi^T \therefore \phi \phi^T = I$

let Λ be a diagonal matrix $\Lambda = diag\{\lambda_1\ \lambda_2, \ldots \lambda_n\}$

from $\Sigma \varphi_i = \lambda_i \varphi_i$

$\Rightarrow \Sigma \phi = \phi \Lambda \quad \therefore \boxed{\Sigma = \phi \Lambda \phi^T}$ or $\boxed{\Lambda = \phi^T \Sigma \phi}$

$$\Lambda = \phi^T \Sigma_x \phi = \phi^T X X^T \phi \overset{Y = \phi^T X}{=\!=} Y Y^T = \Sigma_y = \Lambda$$

Problem of PCA:

① the lost info less important for representing data could be critical for discrimating

② PCA projects data vertially instead of horizontaly.

while maximizes the variation

keeps irrelavent info to discriminate.

EE7403 LEC12.

5. Linear Discriminant Analysis (LDA)

propeties to determine:

maximize saparation between projected class mean

minimize projected with-in class scatter (variance)

$y = \phi^T (x - \mu_i)$

Given q n-dim samples of classes c

$x_1, x_2, \dots x_q$

the num of samples in class w_j is q_j, $j = 1, 2, \dots c$.

$\Sigma_j = \frac{1}{q_j} \Sigma_{x_i \in w_j} (x_i - \mu_j)(x_i - \mu_j)^T$, where $\mu_j = \frac{1}{q_j} \sum_{x_i \in w_j} x_i$

with-class Scatter matrix is:

$$S^w = \sum_{j=1}^{c} \frac{q_j}{q} \Sigma_j$$

between-class scatter matrix

$$S^b = \Sigma_{j=1}^{c} \frac{q_j}{q} (\mu_j - \mu)(\mu_j - \mu)^T \qquad \mu = \frac{1}{q} \sum_{i=1}^{q} x_i = \sum_{j=1}^{c} \frac{q_j}{q} \mu_j$$

LDA: min trace $[\phi^T S^w \phi]$ max trace $[\phi^T S^b \phi]$

$\Rightarrow$ max trace $[\phi^T S^{w-1} S^b \phi] = \sum_{k=1}^{m} \lambda_k^{b/w}$

$$S^t = S^w + S^b$$

Obviously, LDA extracts much more discriminative features than PCA

the rank of S^w is min $[q-c, n]$ at most. (mostly $q - c \ll n$)

Problems: S^w is singular and its inverse does not exist.