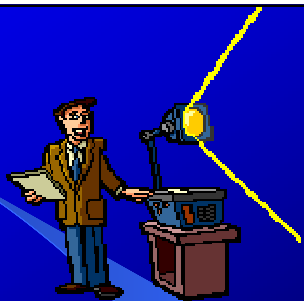
Biometric Authentication

Lecture 5

Biometrics Technology: Pattern Recognition (2)



Outline

- Three PR Approaches
- Principal Component Analysis (PCA)
- Linear Discrimination Analysis (LDA)

Three PR Approaches

PR Approaches

- No single theory of PR can possibly cope with such a broad range of problems. PR applications come in many forms. A given problem may allow one or more of the different solution approaches. One of the guideline principles is just :'Use the right tool for the job!'
- Three typical PR approaches:
  1. Statistical PR (StatPR): There is an underlying and quantifiable statistical basis for the generation of patterns.
  2. Syntactic PR(SyatPR): The underlying structure of the pattern provides the information fundamental for PR.
  3. Neural Networks (NN): Neither of the above cases hold true, but we are able to develop and 'train' an architecture to correctly associate input patterns with desired responses.

Statistical PR Approach (StatPR)

- StatPR - Statistical (or 'decision-theoretic') PR – is a very popular approach in PR. So far, most of PR systems are based on this approach.
- StatPR assumes a statistical basis for classification of algorithms. A set of characteristic measurements, denoted features, are extracted from the input data and are used to assign each feature vector to one of the *c* classes. Features are assumed generated by a state of nature, and therefore the underlying model is of a state of nature or class-conditioned set of probabilities and/or probability density functions.

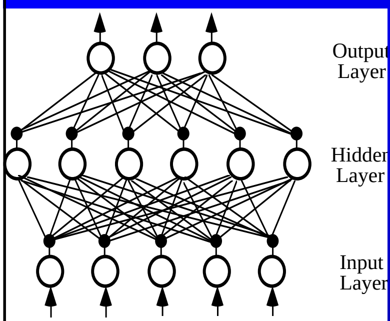
Syntactic PR Approach (SyntPR)

- Many times the significant information in a pattern is not merely in the presence or absence, or the numerical values, of a set of features. Rather , the interrelationships or interconnections of features yield important structural information, which facilitates structural description or classification. This is the basis of syntactic (or structural) PR. However, in using SyntPR approaches, we must be able to quantify and extract structural information and to assess structural similarity of patterns.
- Typically, SyntPR approaches formulate hierarchical descriptions of complex patterns built up from simpler sub-patterns. At the lowest level, primitive elements or 'building blocks' are extracted from the input data.

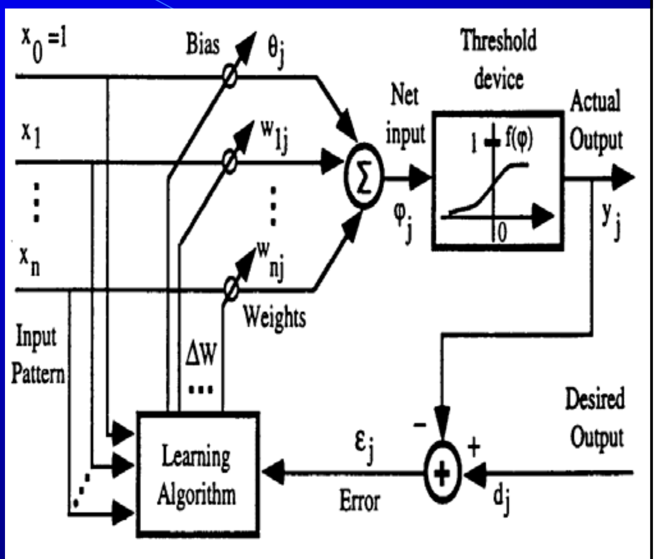
Neural Networks Approach (NN)

- NN is motivated by a desire to try both to understand the brain and to emulate some of its strengths. It can be characterized by
  - ∠ its pattern of connections (architectures)
  - ∠ its method of determining the weights (training or learning)
  - ∠ its activation function
- Developed as generalizations of mathematical models of human cognition or neural biology, based on the assumptions that
  - # Information processing occurs at neurons
  - # Signals are passed between neurons
  - # Each connection link has an associated weight
  - # An activation function is applied to each neurons

Neural Networks Approach (NN)

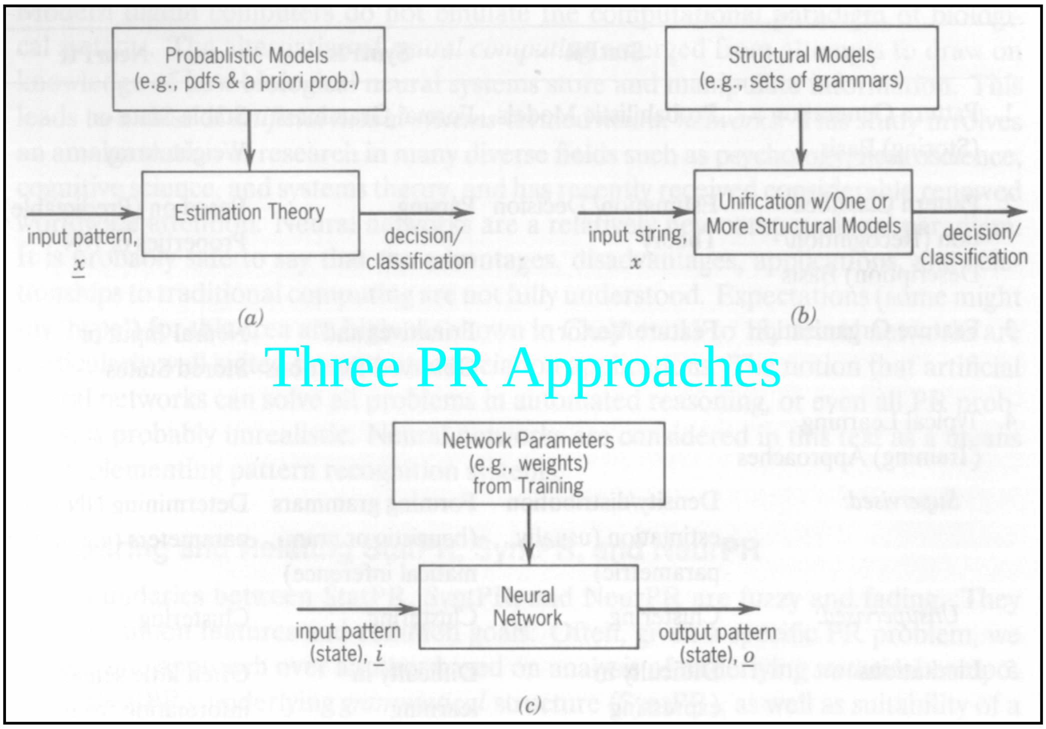


A three-layer NN

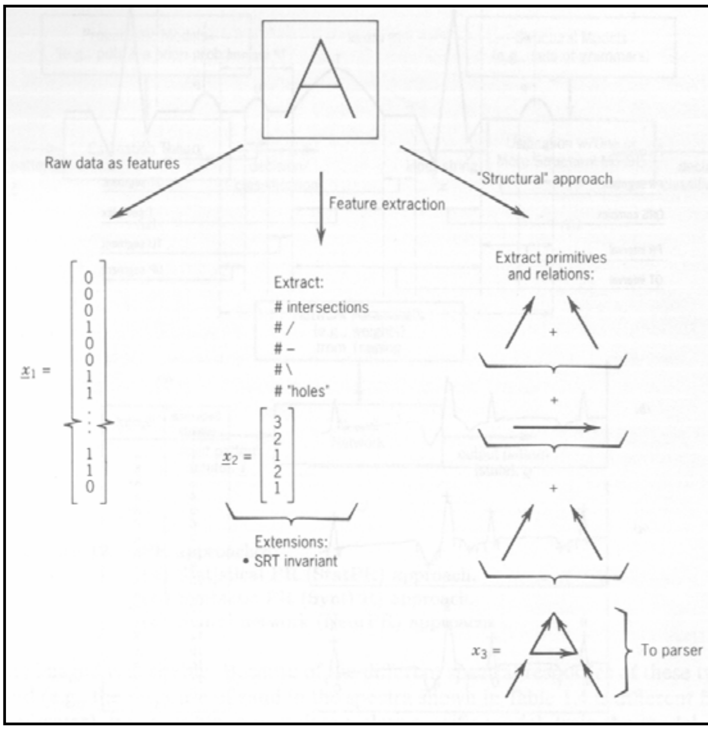


A typical neuron model

Three PR Approaches



Example of PR Approaches



Design Procedure for PR System

- <u>Step 1</u> *Study the classes of patterns under consideration to develop possible characterizations*. This includes assessments of (quantifiable) pattern structure and probabilistic characterizations, as well as exploration of possible within-class and interclass similarity/dissimilarity measures. In addition, possible pattern deformations or invariant properties and characterization of 'noise' sources should be considered at this point.
- <u>Step 2</u> *Determine the availability of feature/ measurement data*.
- <u>Step 3</u> *Consider constraints on desired system performance and computational resources* (e.g.,parts/minutes, classification accuracy).

Design Procedure for PR System

- <u>Step 4</u> *Consider the availability of training data.*
- <u>Step 5</u> *Consider the availability of suitable and known PR techniques* (e.g., StatPR, SyntPR, and NN).
- <u>Step 6</u> *Develop a PR system simulation*. This may involve choosing models, grammars, or network structures.
- <u>Step 7</u> *Train the system*.
- <u>Step 8</u> *Simulate system performance*.
- <u>Step 9</u> *Iterate among the above steps until desired performance is achieved*.

Two Commonly Used StatPR Techniques: PCA & LDA

- PCA – Principal Component Analysis.

  PCA projects the input data into a group of orthogonal coordinate axes, and each data is expressed by a projective vector which can be used in pattern classification. Those axes are called the principal components of the data.

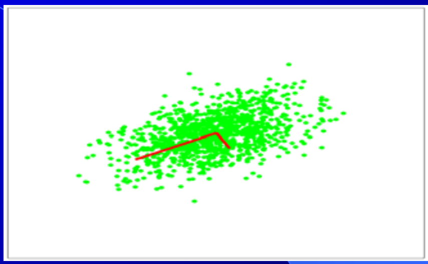
- LDA – Linear Discrimination Analysis.

  LDA applies the famous Fisher criterion to extract the useful pattern discrimination information.

Principal Component Analysis (PCA)

Introduction to PCA

PCA, sometimes called the K-L transformation is a widely used method for reducing the number of dimensions of a data set, which first proposed in 1901 by Pearson. In 1935, Hotelling gave a practical computing method.

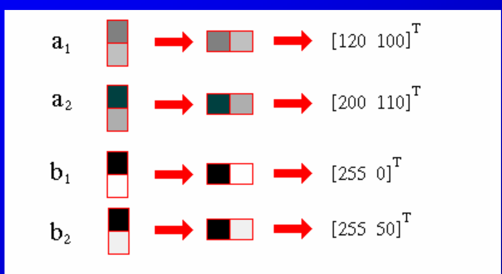


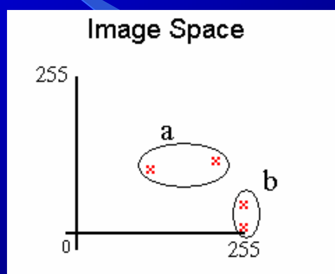
The goal of dimension reduction is to do further processing (feature extraction) or visualization. PCA finds basis vectors for a (sub)space which:

- Maximize the variance retained in the projected data;
- Give uncorrelated projected distributions;
- Minimize the least square reconstruction error.

PCA in Image Space

- Similarly the following 1x2 pixel images are converted into the vectors shown.





- Each image occupies a different point in image space.
- Similar images are near each other in image space.
- Different images are far from each other in image space.

Image Subspace Constructed by PCA

- Principal component analysis is used to calculate the vectors which best represent this small region of image space.
- These are the eigenvectors of the covariance matrix for the training set.
- The eigenvectors are used to define the subspace of images, known as image space.
- Dimensionality reduction procedure used here is called *Karhunen-Loéve transformation* or *principal component analysis.*

Applying the Principal to Biometrics

- A *M* x *N* (e.g., 256x256) pixel image occupies a single point in 65,536-dimensional image space.
- The same type of images occupy a small region of this large image space.
- Similarly, different types should occupy different areas of this smaller region.
- We can identify a person by finding the nearest 'known' vector in image space.

PCA Method (using image data)

Step 1: Suppose that an image data is composed by 2 classes (c1 and c2), each class 10 image samples. Every sample is 28*23 size. Divide this data into training and test sample sets. The training set X is composed by the first 5 samples for each class and the rest constructs the test set Y.

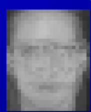
Training set X:

Test set Y:

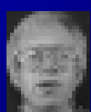
PCA Method (using image data)

Step 2: Express every image sample in X, $X_j^{(i)} \ (i=1,2\,;\ j=1,\cdots,5)$, by a 644 (=28*23) dimensional feature vector. For X, assume that its total mean value is m, the mean values of c1 and c2 are m1 and m2, *E is mathematical expectation* we have:

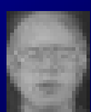
$$m_1 = E\left(X^{(c_1)}\right) = \sum_{j=1}^{5} p_j^{(1)} X_j^{(1)} = \sum_{j=1}^{5} \frac{1}{5} X_j^{(1)}$$ ->



$$m_2 = E\left(X^{(c_2)}\right) = \sum_{j=1}^{5} p_j^{(2)} X_j^{(2)} = \sum_{j=1}^{5} \frac{1}{5} X_j^{(2)}$$ ->



$$m = E(m_1, m_2) = \sum_{i=1}^{2} p^{(i)} m_i = \sum_{i=1}^{2} \frac{1}{2} m_i$$ ->



PCA Method (using image data)

Step 3: Compute the total scatter matrix S_t :

$$S_t = E\left[(X-m)(X-m)^T\right]$$

$$= \sum_{i=1}^{2} p_i \left[\sum_{j=1}^{5} p_j^{(i)} \left(X_j^{(i)} - m\right)\left(X_j^{(i)} - m\right)^T\right]$$

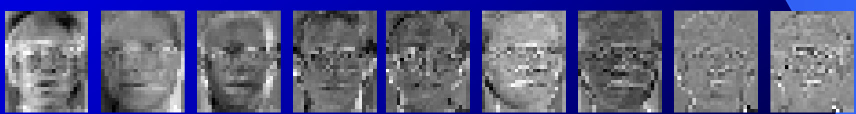
$$= \sum_{i=1}^{2} \frac{1}{2} \left[\sum_{j=1}^{5} \frac{1}{5} \left(X_j^{(i)} - m\right)\left(X_j^{(i)} - m\right)^T\right]$$

PCA Method (using image data)

Step 4: Compute the eigenvalue and eigenvectors of S_t We have

$$\lambda_i \varphi_i = S_t \varphi_i \quad , \quad i = 1, \cdots, 644 \quad ,$$

where λ_i and φ_i are the i^{th} eigenvalue and eigenvector. Each vector is 644 dimensional. We re-express 9 eigenvectors with nonzero eigenvalues in the form of image as following (they are also called Eigenfaces):



Select some eigenvectors that have the largest eigenvalues to represent the most significant variation within the image set

PCA Method (using image data)

Step 5: Select the most principal components or eigenvectors. From the following figure, we find that

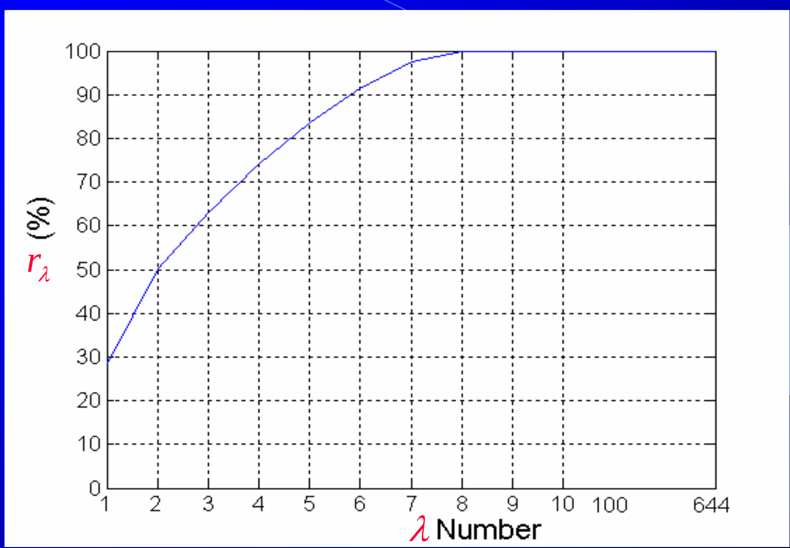
$$\lambda_{10} = \lambda_{11} = \cdots = \lambda_{644} = 0 \quad ,$$

and

$$r_{\lambda} = \frac{\sum_{i=1}^{6} \lambda_i}{\sum_{j=1}^{644} \lambda_j} \times 100\% = 91.32\% > 90\% \quad ,$$

where r_{λ} is the ratio of the eigenvalue sum of selected components to the total sum (refer to the following figure).

PCA Method (using image data)



PCA Method (using image data)

Step 6: The PCA projection transform W is composed by:

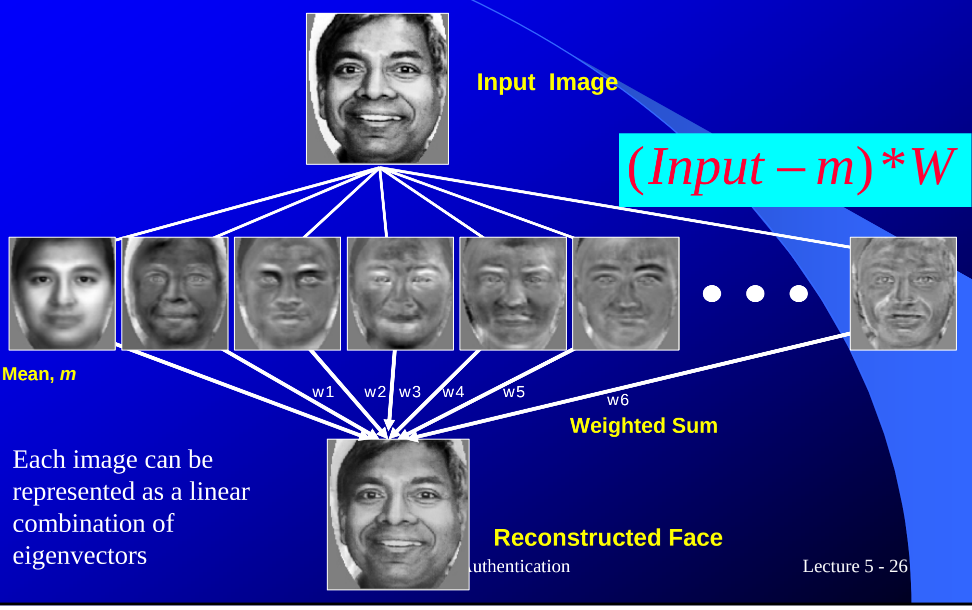
$$W = (\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6)$$

Then, for X and Y, we obtain their transformed feature sets X' and Y' :

$$X' = (X - m) * W, \quad Y' = (Y - m) * W,$$

Notice: For a given input image, its new feature vector can be formed by only has 6 dimension. Then, we can apply any classifier to do the classification, e.g. nearest neighbor classifier.

Face Reconstruction Demo



Each image can be represented as a linear combination of eigenvectors

Eigenface in Practice

- Align a set of face images (the training set)
  - Rotate, scale and translate such that the eyes are located at the same coordinates.
- Compute the average face image
- Compute the difference image for each image in the training set
- Compute the covariance matrix of this set of difference images
- Compute the eigenvectors of the covariance matrix

Examples of Eigenfaces

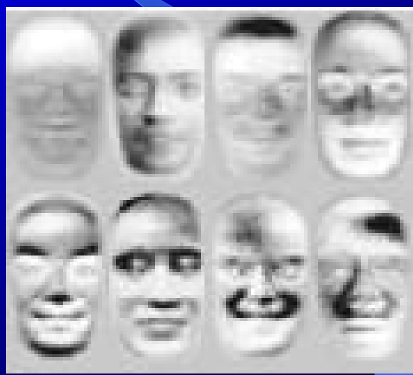
- The eigenvectors of the covariance matrix can be viewed as images.

These are the first 4 eigenvectors, from a training set of 23 images....

Hence the name eigenfaces.

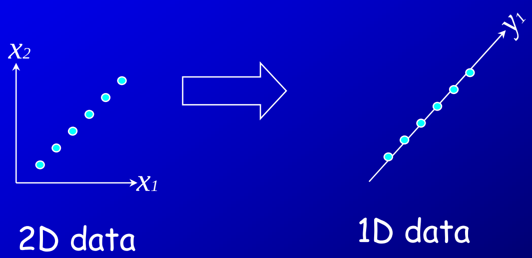
Eigenfaces

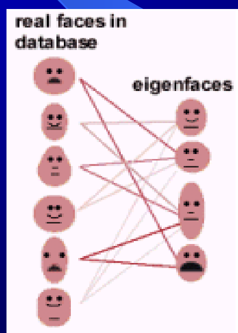
Principal components are called “eigenfaces” and they span the “face space”



Dimensionality Reduction

- Only selecting the top *M* eigenvectors, reduces the dimensionality of the data.
- Too few eigenvectors results in too much information loss, and hence less discrimination between images.



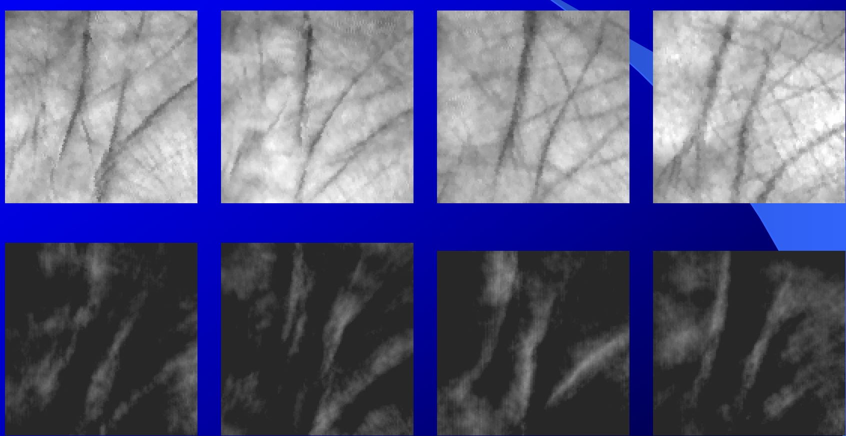


PCA Processing Procedure for Image Data

- Training:
  - Acquire initial set of images (training set)
  - Calculate the eigenfaces from the training set, keeping only the *M* images corresponding to the highest eigenvalues
  - Calculate representation of each known individual in given space
- Testing:
  - Project input image into image space
  - Find most likely candidate by distance computation

Other PCA Application: Eigenpalms

Sub-images of the palmprint are analyzed by eigenpalm



Eigenfaces Problems

- Recognition performance decreases quickly as the head size, or scale, is misjudged. The head size in the input image must be close to that of the eigenfaces for the system to work well
- In the case where every face image is classified as known, a sample system achieved approximately 96% correct classification averaged over lighting variation, 85% correct averaged over orientation variation, and 64% correct averaged over size variation

PCA Summary

Advantages

1. Reduce data dimensionality.
2. Satisfy the Minimal MSE (Mean Square Error) Rule.
3. Eliminate the correlation of original data.

Disadvantages

- Eigenfaces do not distinguish between shape and appearance
- PCA does not use class information :

  PCA projections are optimal for reconstruction from a low dimensional basis, they may not be optimal from a discrimination standpoint: "Much of the variation from one image to the next is due to illumination changes." [Moses, Adini, Ullman]

Linear Discrimination Analysis (LDA)

LDA Introduction (Using Image Data)

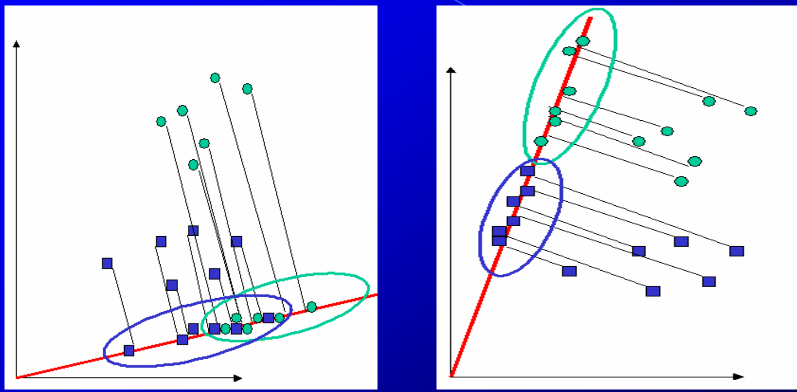
Step 1: For vector data X (see *Step 1 & 2* in PCA), compute within-class scatter matrix S_w and between-class scatter matrix S_b ,

$$S_w = \sum_{i=1}^{c} P_i E\left[(X - m_i)(X - m_i)^T \right] \quad , \quad S_b = \sum_{i=1}^{c} P_i \left[(m_i - m)(m_i - m)^T \right],$$

where, P_i and m_i are the prior probability and the mean value of the *i-th* class.

Step 2: Compute the eigenvectors of $S_w^{-1} S_b$, which construct the LDA transform.

Fisher's Linear Discriminant



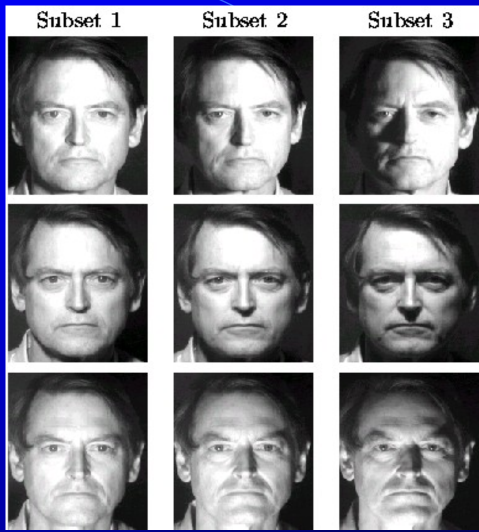
- Attempts to project the data such that the classes are separated.

LDA Advantages

1. LDA applies the Fisher criterion, which is not only simple, but also very efficient in discriminative feature extraction for the classification. This has been proved by many pattern recognition experiments.

2. LDA can further reduce the feature dimension for pattern classification compared with PCA, since the rank of $S_w^{-1} S_b$ is generally equal to *c - 1.*

Application: Fisherfaces



Application: Fisherfaces

Method	Subset 1	Subset 2	Subset 3
Eigenface	0.0	4.4	41.5
Eigenface w/o 1^{st} 3	0.0	4.4	27.7
Fisherface	0.0	0.0	4.6

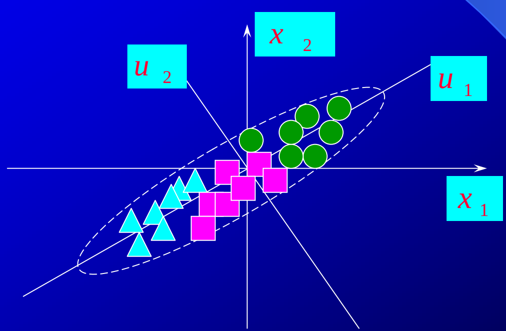
Other Application: Fisherpalm

(a) (b) (c)

An example of the Fisherpalm in the case of two palmprint classes: (a) and (b) are samples in the class one and two, respectively, and (c) is the Fisherpalm used for classification

Demo of PCA Transform

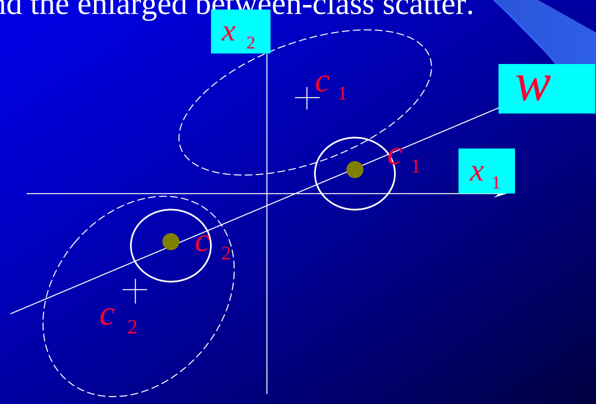
There are 3-class samples with two-dimensional features (x_1 and x_2). By PCA, two components (u_1 and u_2) are obtained. And, we can remove u_2 to reduce feature dimension.



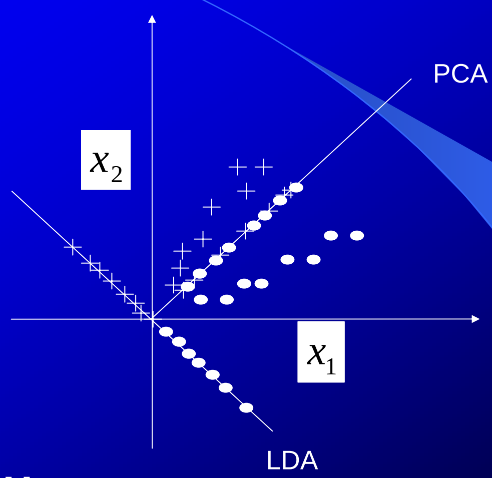
Principal components are the eigenvectors of the covariance matrix of the set of images

Demo of LDA Transform

There are 2-class samples with two features (x_1 and x_2). By LDA, an optimal projection direction w is obtained. The solid circles of c_1 and c_2 denote the compressed within-class scatter and the enlarged between-class scatter.



Comparison between PCA & LDA



A two-class problem

Eigenface & Fisherface

- *Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection*
  - P. Belhumeur, J. Hespanha, D. Kriegman
    - Yale University
- Eigenfaces attempt to maximise the scatter of the training images in face space.
- Fisherfaces attempt to maximise the between class scatter, while minimising the within class scatter.
- In other words, moves images of the same face closer together, while moving images of difference faces further apart.

Questions?

- There are three PR approaches: StatPR, SyatPR and NN. What difference between them? Based on your knowledge, can you give a simple application for each approach?
- What is the main stages about PCA? Which difference between PCA and LDA?
- Can you find some PCA problems? How to improve PCA/LDA performance?

So much for today!

Thank you !!!