

Tutorial 4

Qu Xiaofeng, Teaching Assistant
quxiaofeng.at.polyu@gmail.com

COMP435p
Biometrics Authentication

February 13, 2012





Table of Contents

1 Problems

- Problem 1: Answer the questions
- Problem 2: Three PR Approaches
- Problem 3: StatPR and SyatPR
- Problem 4: PCA
- Problem 5: Three PR Approaches



Outline

1 Problems

- Problem 1: Answer the questions
- Problem 2: Three PR Approaches
- Problem 3: StatPR and SyatPR
- Problem 4: PCA
- Problem 5: Three PR Approaches

Problem 1.1 scatter matrix, eigenvector and eigenvalue



What is the scatter matrix (P5:21)? Understand what about eigenvector and eigenvalue as well as their functions? (P5:22)

Problem 1.1 scatter matrix, eigenvector and eigenvalue



PCA Method (using image data)

Step 3: Compute the total scatter matrix S_t :

$$\begin{aligned} S_t &= E[(X-m)(X-m)^T] \\ &= \sum_{i=1}^2 P_i \left[\sum_{j=1}^5 p_j^{(i)} (X_j^{(i)} - m)(X_j^{(i)} - m)^T \right] \\ &= \sum_{i=1}^2 \frac{1}{2} \left[\sum_{j=1}^5 \frac{1}{5} (X_j^{(i)} - m)(X_j^{(i)} - m)^T \right] \end{aligned}$$

Problem 1.1 scatter matrix, eigenvector and eigenvalue

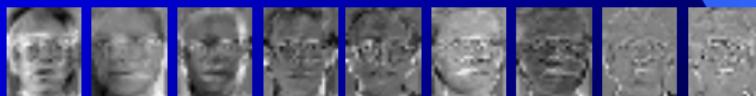


PCA Method (using image data)

Step 4: Compute the eigenvalue and eigenvectors of S_I . We have

$$\lambda_i \varphi_i = S_I \varphi_i \quad , \quad i = 1, \dots, 644 \quad ,$$

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

Problem 1.1 scatter matrix, eigenvector and eigenvalue



"Detailed links(click)"

"PCA books(click)"

kolho3.tiera.ru



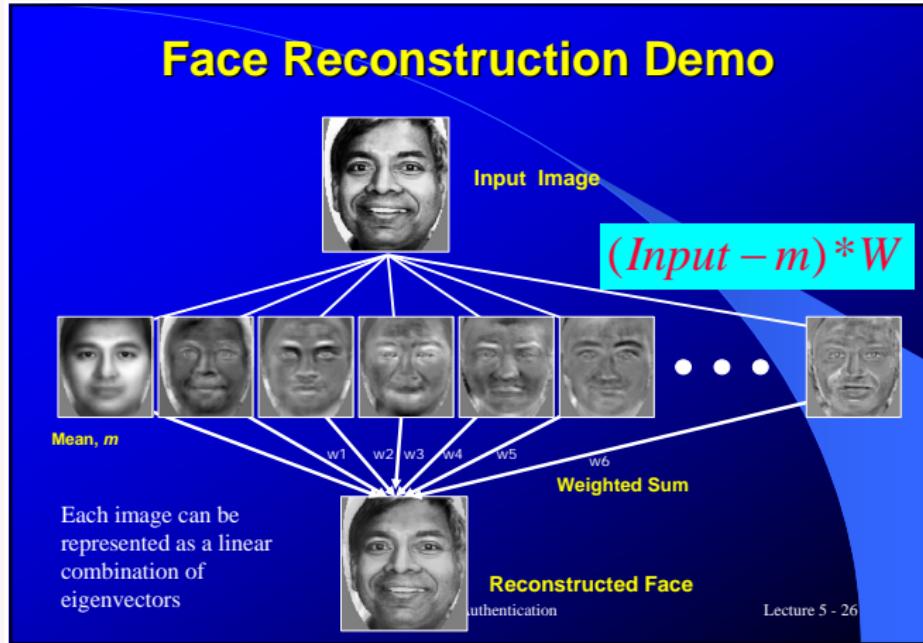


Problem 1.2 Eigenface

Understand the PCA application to facial recognition: Eigenface. (P5:26-30)



Problem 1.2 Eigenface





Problem 1.2 Eigenface

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





Problem 1.2 Eigenface

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.



Problem 1.2 Eigenface

Eigenfaces

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



Biometric Authentication

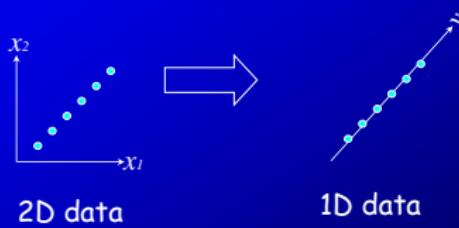
Lecture 5 - 29



Problem 1.2 Eigenface

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.





Problem 1.3 PCA and LDA

Compare two StatPR techniques, PCA and LDA (P5:13) and point out their main difference (P5:42-45)



Problem 1.3 PCA and LDA

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.

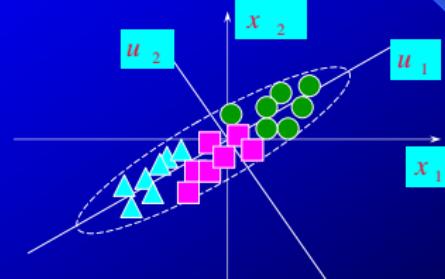
LDA applies the famous Fisher criterion to extract the useful pattern discrimination information.



Problem 1.3 PCA and LDA

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.



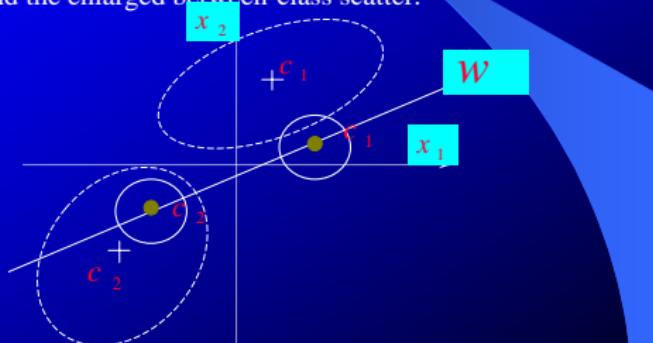
Principal components are the eigenvectors of the covariance matrix of the set of images



Problem 1.3 PCA and LDA

Demo of LDA Transform

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

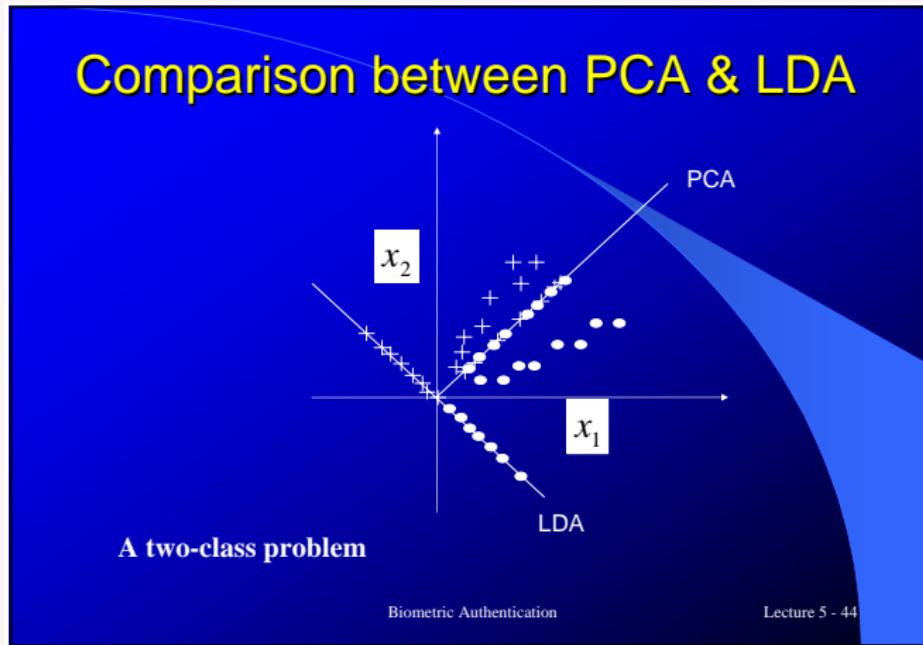


Biometric Authentication

Lecture 5 - 43



Problem 1.3 PCA and LDA





Problem 1.3 PCA and LDA

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 different faces further apart.





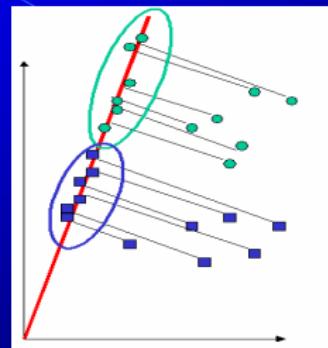
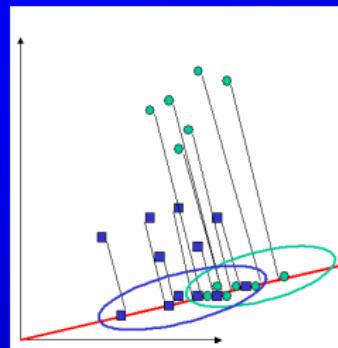
Problem 1.4 LDA

Linear discrimination analysis (LDA) is introduced in P5:37-40. Please understand the two steps in P5:37 and compare within-class scatter matrix with between-class scatter matrix.



Problem 1.4 LDA

Fisher's Linear Discriminant



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



Problem 1.4 LDA

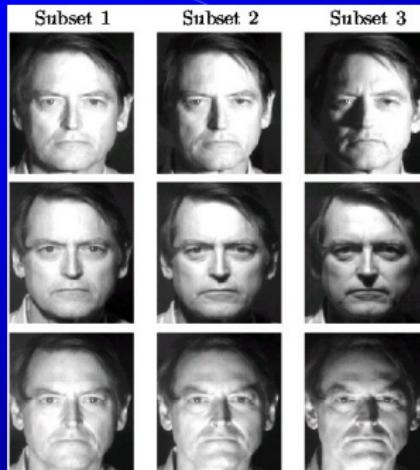
LDA Advantages

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 $c - 1$.



Problem 1.4 LDA

Application: Fisherfaces



Biometric Authentication

Lecture 5 - 39



Problem 1.4 LDA

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





Outline

1 Problems

- Problem 1: Answer the questions
- **Problem 2: Three PR Approaches**
- Problem 3: StatPR and SyatPR
- Problem 4: PCA
- Problem 5: Three PR Approaches



Problem 2: Three PR Approaches

There are three PR approaches: StatPR, SyatPR and NN. What difference between them (P5:4-10)? Based on your knowledge, can you give a simple application for each approach?



Problem 2: 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.



Problem 2: Three PR Approaches

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**.



Problem 2: Three PR Approaches

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.



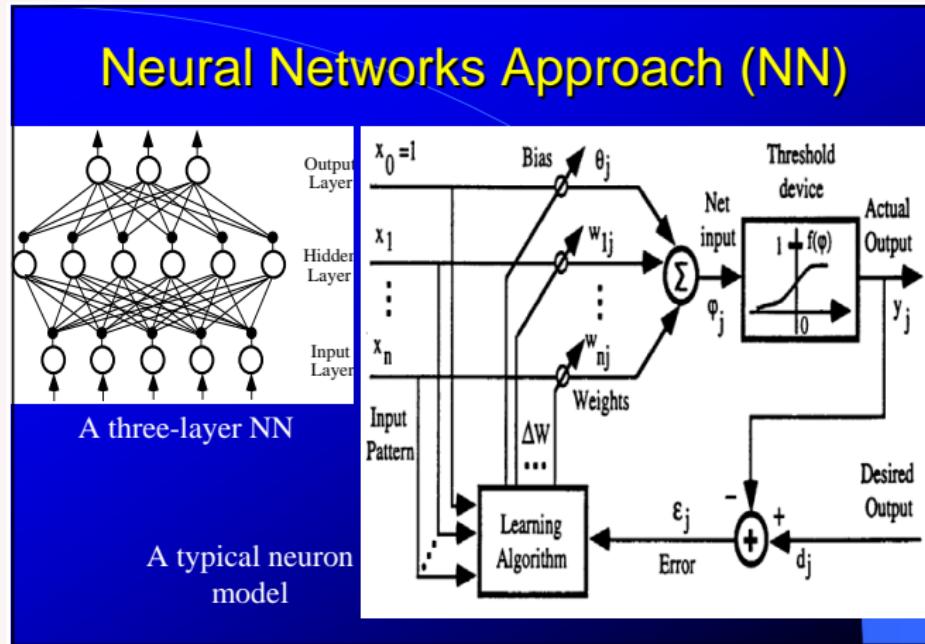
Problem 2: Three PR Approaches

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

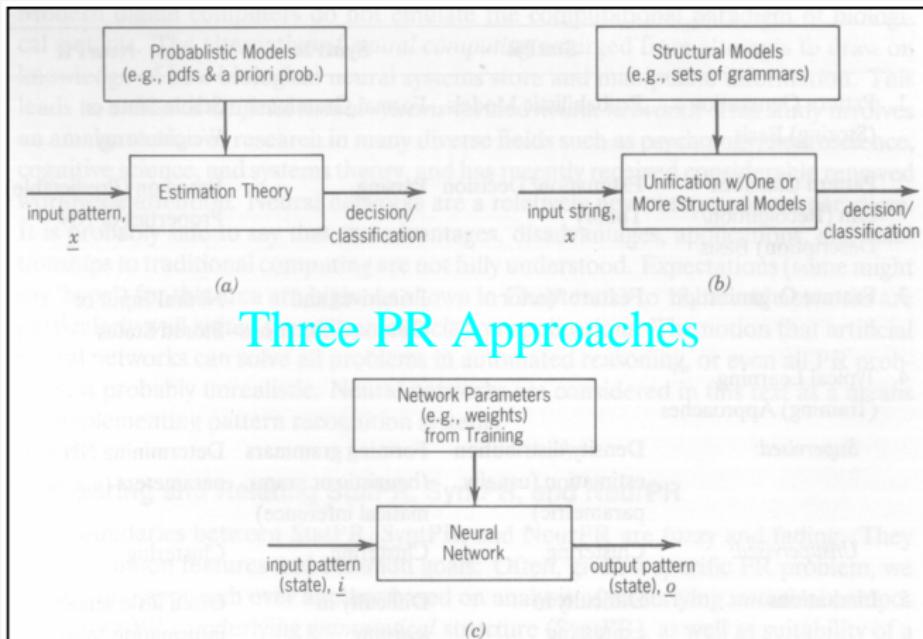


Problem 2: Three PR Approaches



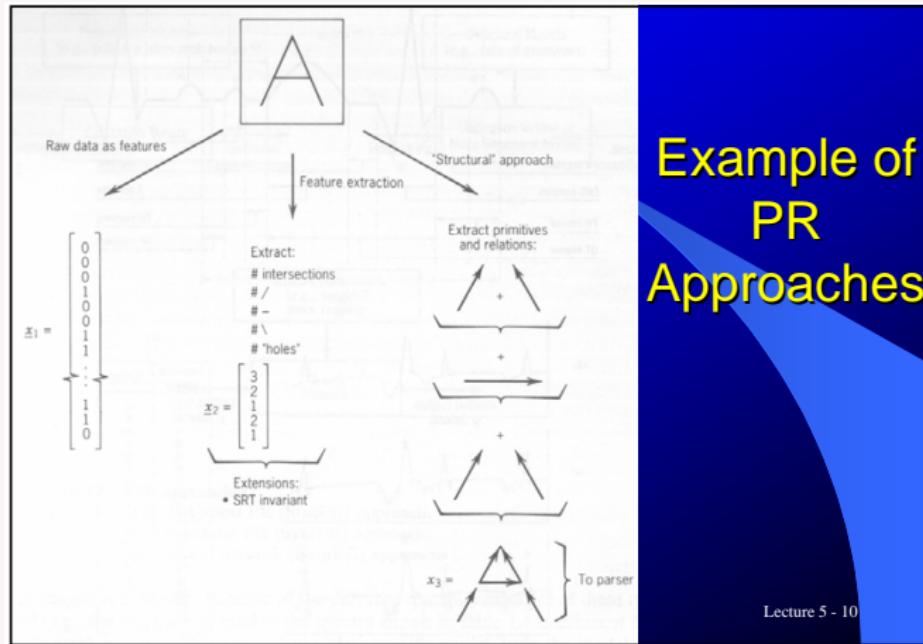


Problem 2: Three PR Approaches





Problem 2: Three PR Approaches





Problem 2: Three PR Approaches

- The properties of the patterns (entities) are different:
- For statistical PR the patterns can be extracted a set of features and there is an underlying statistical model for the generation of these patterns.
- For syntactic PR the patterns should contain structural or relational information and the structure of an entity is paramount.
- For NN, the patterns are stored in the Network, and it can't be seen.



Problem 2: Three PR Approaches

- The methods are different:
- For statistical PR the problems are solved by using statistical theory.
- For syntactic PR the classification can be accomplished by defining suitable and distinct grammars that reflect the structure of each pattern class.
- For NN, the result is learned by the neurons in the network.



Problem 2: Three PR Approaches

Examples

Approaches	Examples
StatPR	Face, Palmprint
SyatPR	Disease
NN	Motor Control





Outline

1 Problems

- Problem 1: Answer the questions
- Problem 2: Three PR Approaches
- **Problem 3: StatPR and SyatPR**
- Problem 4: PCA
- Problem 5: Three PR Approaches

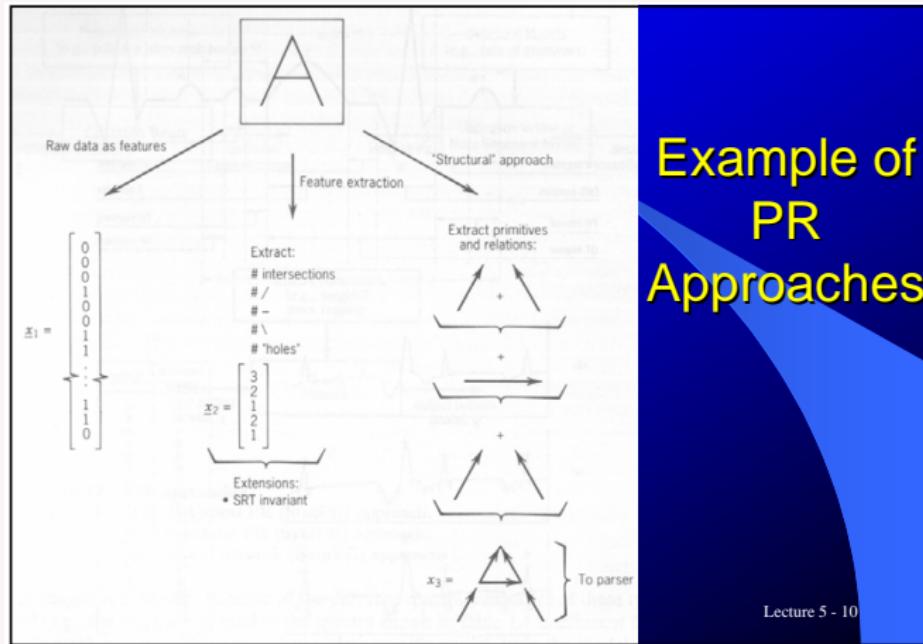


Problem 3: StatPR and SyatPR

Please check the examples of PR approaches in P5:10. Try to analyze the character "H" by statistical PR approach and structural PR approach.



Problem 3: StatPR and SyatPR





Problem 3: StatPR and SyatPR

StatPR approach

The feature set:(intersections, -, |, holes)

$$\mathbf{x} = [2, 1, 4, 0]$$

SyntPR approach

Primitives and relations: $\uparrow + \uparrow \rightarrow \uparrow + \uparrow$





Outline

1 Problems

- Problem 1: Answer the questions
- Problem 2: Three PR Approaches
- Problem 3: StatPR and SyatPR
- **Problem 4: PCA**
- Problem 5: Three PR Approaches



Problem 4: PCA

From P5:19-25, PCA method given by using image data is defined, which projects an image space with 644 dimension into 6 dimension eigenvector space. Please understand each step.



Problem 4: PCA

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**:





Problem 4: PCA

PCA Method (using image data)

Step 2: Express every image sample in X , $X_j^{(i)} (i=1,2; j=1,\dots,5)$,

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:

$$m_1 = E(X^{(c_1)}) = \sum_{j=1}^5 p_j^{(1)} X_j^{(1)} = \sum_{j=1}^5 \frac{1}{5} X_j^{(1)} \rightarrow \begin{array}{c} \text{Portrait of a man} \\ \text{with glasses} \end{array}$$

$$m_2 = E(X^{(c_2)}) = \sum_{j=1}^5 p_j^{(2)} X_j^{(2)} = \sum_{j=1}^5 \frac{1}{5} X_j^{(2)} \rightarrow \begin{array}{c} \text{Portrait of a man} \\ \text{without glasses} \end{array}$$

$$m = E(m_1, m_2) = \sum_{i=1}^2 p_i^{(i)} m_i = \sum_{i=1}^2 \frac{1}{2} m_i \rightarrow \begin{array}{c} \text{Average portrait} \\ \text{of a man with} \\ \text{and without glasses} \end{array}$$



Problem 4: PCA

PCA Method (using image data)

Step 3: Compute the total scatter matrix S_t :

$$\begin{aligned} S_t &= E[(X-m)(X-m)^T] \\ &= \sum_{i=1}^2 p_i \left[\sum_{j=1}^5 p_j^{(i)} (X_j^{(i)} - m)(X_j^{(i)} - m)^T \right] \\ &= \sum_{i=1}^2 \frac{1}{2} \left[\sum_{j=1}^5 \frac{1}{5} (X_j^{(i)} - m)(X_j^{(i)} - m)^T \right] \end{aligned}$$



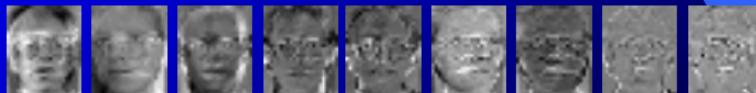
Problem 4: PCA

PCA Method (using image data)

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

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

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



Problem 4: PCA

PCA Method (using image data)

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

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

and

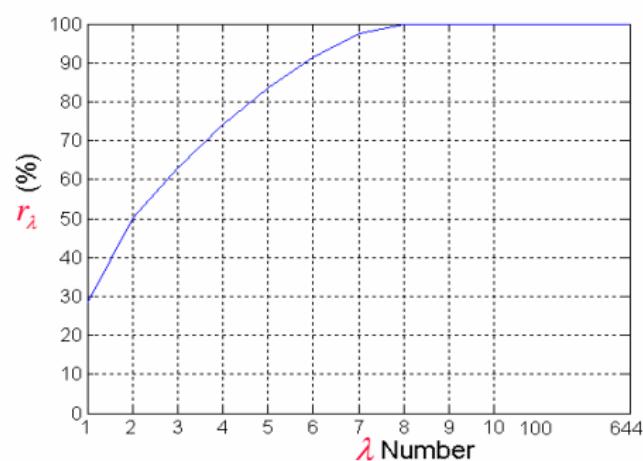
$$r_\lambda = \frac{\sum_{i=1}^6 \lambda_i}{\sum_{j=1}^{644} \lambda_j} \times 100 \% = 91.32 \% > 90 \% \quad ,$$

where r_λ is the ratio of the eigenvalue sum of selected components to the total sum (refer to the following figure).



Problem 4: PCA

PCA Method (using image data)



Biometric Authentication

Lecture 5 - 24



Problem 4: PCA

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 $\overset{\circ}{X}$ and $\overset{\circ}{Y}$:

$$\overset{\circ}{X} = (X - m) * W, \quad \overset{\circ}{Y} = (Y - m) * W,$$

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.





Outline

1 Problems

- Problem 1: Answer the questions
- Problem 2: Three PR Approaches
- Problem 3: StatPR and SyatPR
- Problem 4: PCA
- **Problem 5: Three PR Approaches**

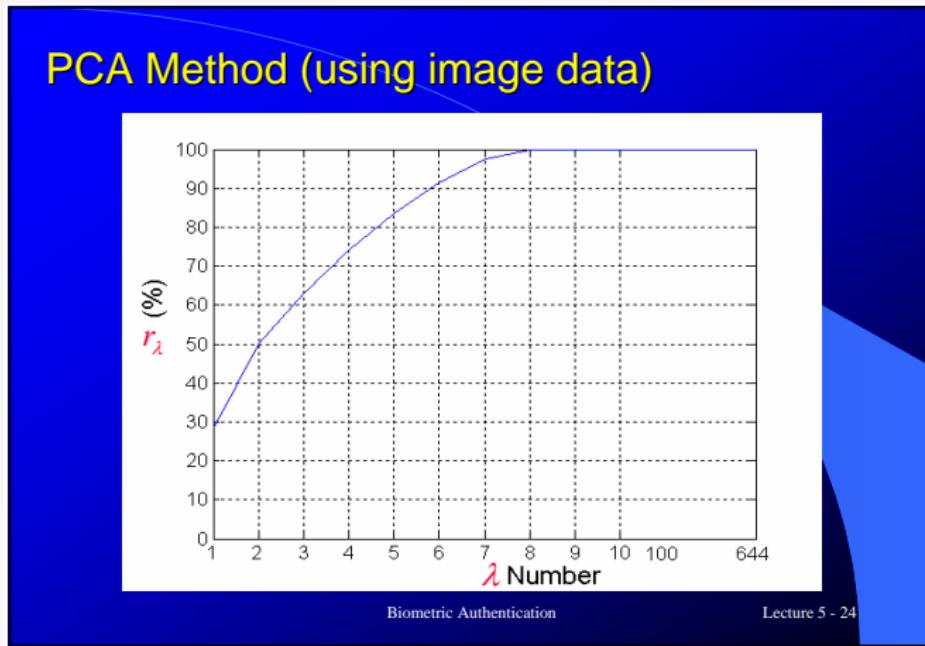


Problem 5: Three PR Approaches

According to the figure in P5:24, how to find its minimum λ if we hope to get $\gamma_\lambda > 60$?



Problem 5: Three PR Approaches



When $\lambda = 2$ the $\gamma_\lambda < 60$ When $\lambda = 3$ the $\gamma_\lambda > 60$, so $\lambda = 3$ is what we want.



Problems

Any questions?