Principal Component Analysis (PCA) : An Overview

Presentation · March 2016

CITATIONS
0

READS
1,201

1 author:

Alaa Tharwat
Frankfurt University of Applied Sciences
78 PUBLICATIONS **539** CITATIONS

SEE PROFILE

Principal Component Analysis: An Overview

Alaa Tharwat

Alaa Tharwat. "Principal component analysis-a tutorial" International Journal of Applied Pattern Recognition 3(3) (2016) 197-240
Email: engalaatharwat@hotmail.com

Feb. 27, 2018

- Introduction to Principal component analysis (PCA)
  - Principal Components (PCs)
  - Covariance Matrix Method
  - Singular Value Decomposition (SVD) Method
  - PCA Space (Lower Dimensional Space)
  - Data Reconstruction
  - PCA Algorithms
- Numerical example
  - First Example: 2D-Class Example
  - Multi-Class Example
- Experimental Results and Discussion
  - Biometric Experiment
  - Image Compression Experiment
  - Data Visualization Experiment
- Conclusions

- Introduction to Principal component analysis (PCA)
  - Principal Components (PCs)
  - Covariance Matrix Method
  - Singular Value Decomposition (SVD) Method
  - PCA Space (Lower Dimensional Space)
  - Data Reconstruction
  - PCA Algorithms
- Numerical example
  - First Example: 2D-Class Example
  - Multi-Class Example
- Experimental Results and Discussion
  - Biometric Experiment
  - Image Compression Experiment
  - Data Visualization Experiment
- Conclusions

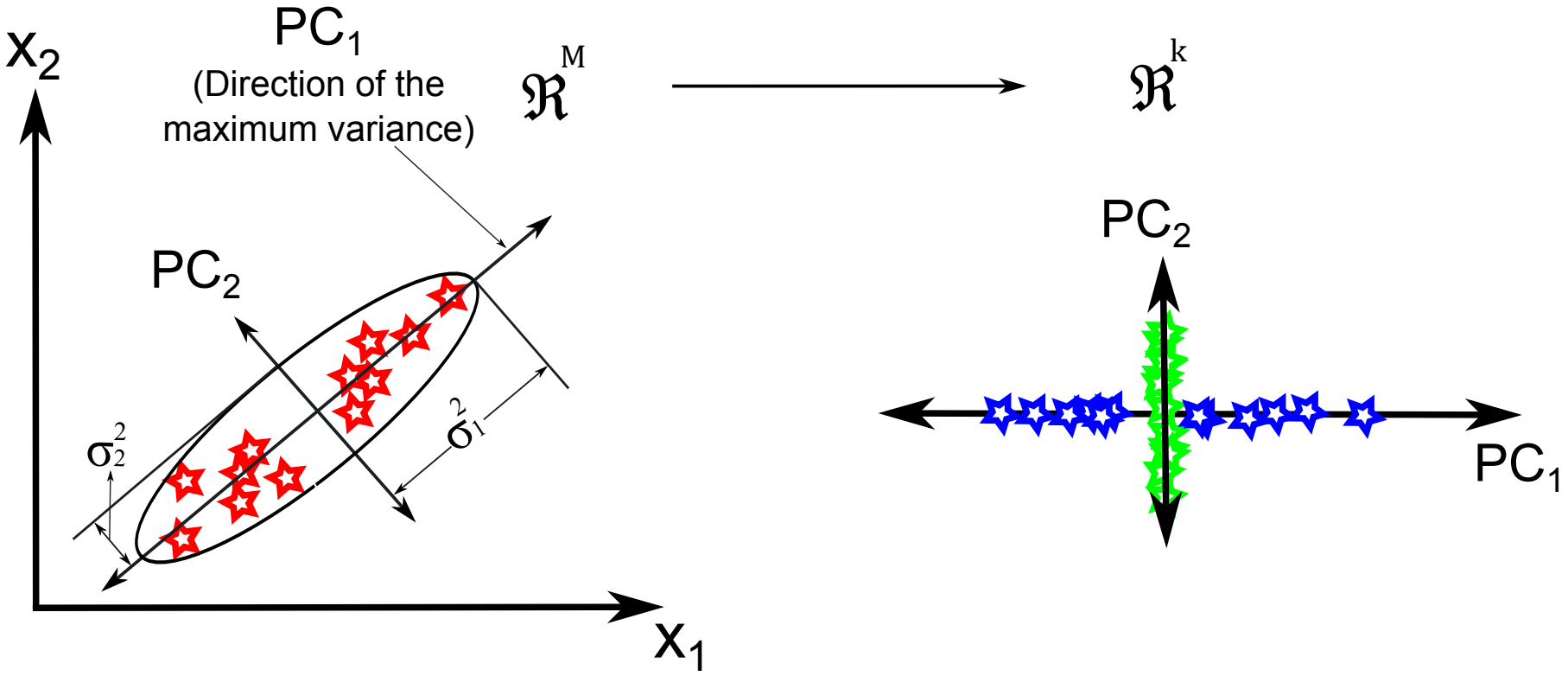
- The goal for any dimensional reduction method is to reduce the dimensions of the original data for different purposes such as visualization, decrease CPU time, ..etc..
- Dimensionality reduction techniques are important in many applications related to machine learning, data mining, Bioinformatics, biometric and information retrieval.
- There are two types of dimensionality reduction methods, namely, supervised and unsupervised.
  - **Supervised** (e.g. LDA).
  - **Unsupervised** (e.g. PCA).

- The goal of the PCA technique is to find a lower dimensional space or PCA space (W) that is used to transform the data ($X = \{x_1, x_2, \ldots, x_N\}$) from a higher dimensional space ($\mathcal{R}^M$) to a lower dimensional space ($\mathcal{R}^k$), where
  - N represents the total number of samples or observations,
  - x_i represents i^{th} sample, pattern, or observation
- All samples have the same dimension ($x_i \in \mathcal{R}^M$). In other words, each sample is represented by M variables, i.e. each sample is represented as a point in M-dimensional space.

- The direction of the PCA space represents the direction of the maximum variance of the given data.
- The PCA space is consists of a number of PCs. Each principal component has a different robustness according to the amount of variance in its direction.
- The figure below shows an example of the two-dimensional data (x_1, x_2). The original data are on the left with the original coordinate, i.e. x_1 and x_2, the variance of each variable is graphically represented and the direction of the maximum variance, i.e. the principal component PC_1, is shown; on the right the original data are projected on the first (blue stars) and second (green stars) principal components.

PCA

Orthogonal Transformation
Axes Rotation



- Introduction to Principal component analysis (PCA)
  - Principal Components (PCs)
  - Covariance Matrix Method
  - Singular Value Decomposition (SVD) Method
  - PCA Space (Lower Dimensional Space)
  - Data Reconstruction
  - PCA Algorithms
- Numerical example
  - First Example: 2D-Class Example
  - Multi-Class Example
- Experimental Results and Discussion
  - Biometric Experiment
  - Image Compression Experiment
  - Data Visualization Experiment
- Conclusions

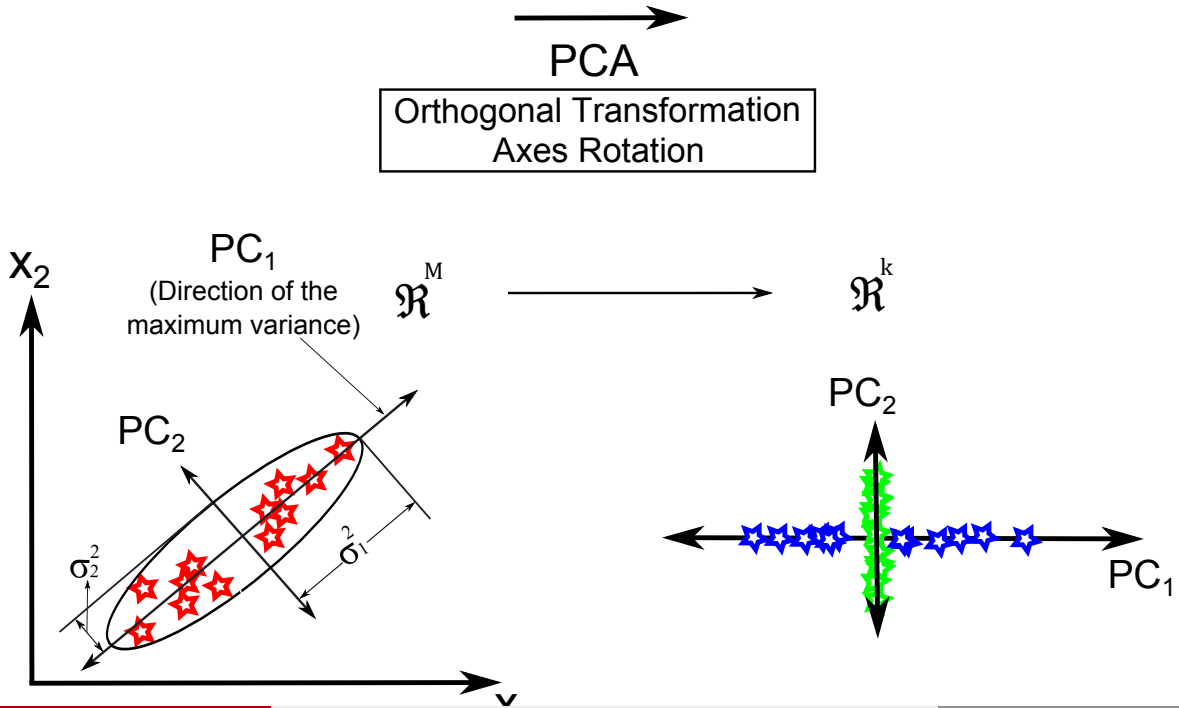
- The PCA space consists of k principal components.
- The principal components are orthonormal[1], uncorrelated[2], and it represents the direction of the maximum variance.
- The first principal component ($(PC_1$ or $v_1) \in \mathcal{R}^{M \times 1}$) of the PCA space represents the direction of the maximum variance of the data, the second principal component has the second largest variance, and so on.

[1] Orthonormal vectors have a unit length and orthogonal as follows,

$$v_i^T v_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

[2] v_i and v_j are uncorrelated if $Cov(v_i, v_j) = 0,\ i \neq j$, where $Cov(v_i, v_j)$ represents the covariance between the i^{th} and j^{th} vectors.

- The figure below shows how the original data are transformed from the original space ($\mathcal{R}^M$) to the PCA space ($\mathcal{R}^k$). Thus, the PCA technique is considered an orthogonal transformation due to its orthogonal principal components or axes rotation due to the rotation of the original axes.
- There are two methods to calculate the principal components. The first method depends on calculating the *covariance matrix*, while, the second one uses the *Singular value decomposition SVD* method.

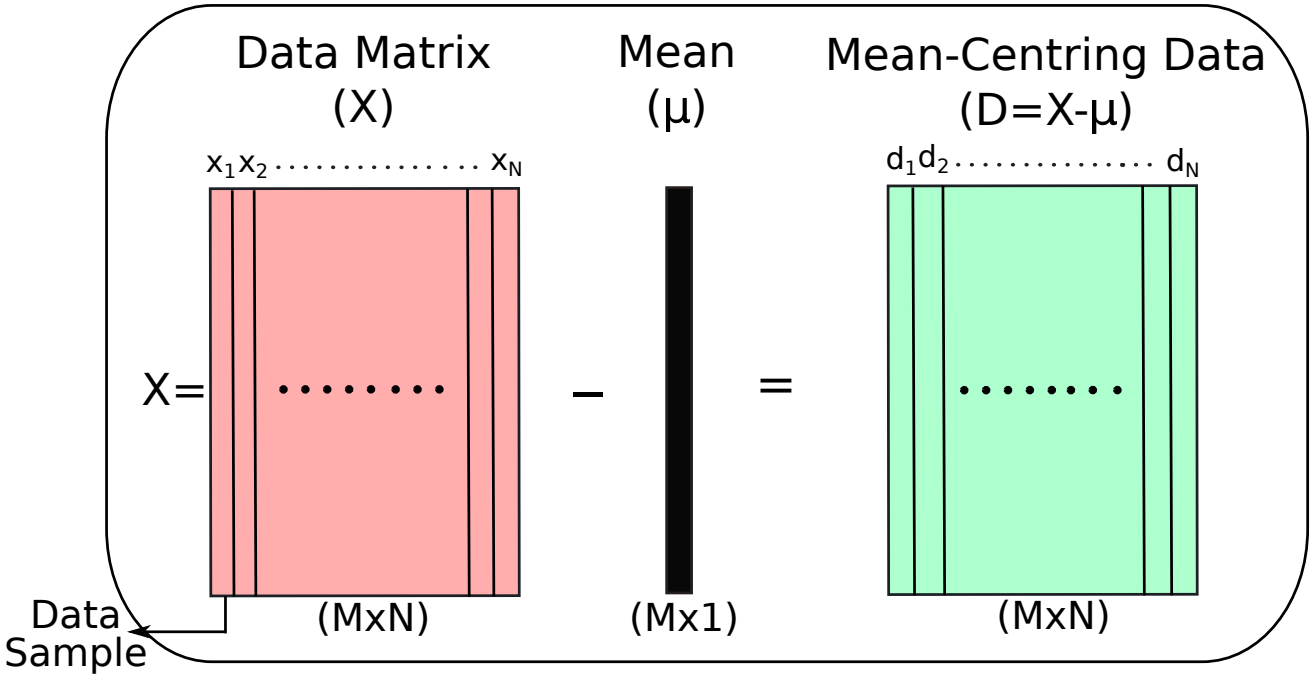


- Introduction to Principal component analysis (PCA)
  - Principal Components (PCs)
  - Covariance Matrix Method
  - Singular Value Decomposition (SVD) Method
  - PCA Space (Lower Dimensional Space)
  - Data Reconstruction
  - PCA Algorithms
- Numerical example
  - First Example: 2D-Class Example
  - Multi-Class Example
- Experimental Results and Discussion
  - Biometric Experiment
  - Image Compression Experiment
  - Data Visualization Experiment
- Conclusions

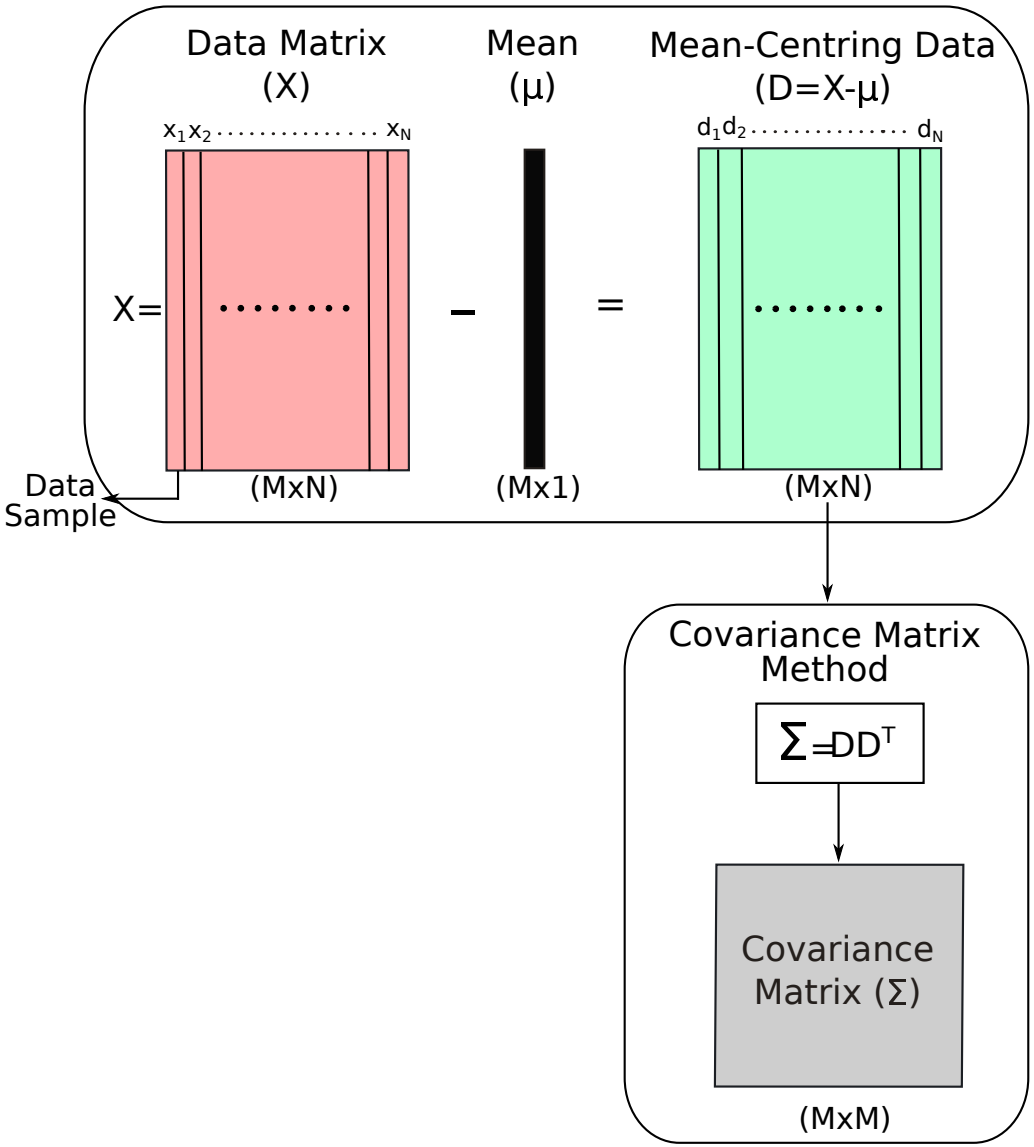
- In this method, there are two main steps to calculate the PCs of the PCA space.
  1. The covariance matrix of the data matrix (X) is calculated.
  2. The eigenvalues and eigenvectors of the covariance matrix are calculated.

- The variance of any variable measures the deviation of that variable from its mean value and it is defined as follows,
  $\sigma^2(x) = Var(x) = E((x-\mu)^2) = E\{x^2\} - (E\{x\})^2$, where μ represents the mean of the variable x, and $E(x)$ represents the expected value of x.
- The covariance matrix is used when the number of variables more than one and it is defined as follows,
  $\Sigma_{ij} = E\{x_i x_j\} - E\{x_i\}E\{x_j\} = E[(x_i - \mu_i)(x_j - \mu_j)]$.

- After calculating the mean of each variable in the data matrix, the mean-centering data are calculated by subtracting the mean ($\mu \in \mathcal{R}^{(M \times 1)}$) from each sample as follows,
$D = \{d_1, d_2, \ldots, d_N\} = \{x_1 - \mu, x_2 - \mu, \ldots, x_N - \mu\}$.



- The covariance matrix is then calculated as follows, $\Sigma = DD^T$.



- Covariance matrix is a symmetric matrix (i.e. $X = X^T$) and always positive semi-definite matrix [3].
- The diagonal values of the covariance matrix represent the variance of the variable $x_i, i = 1, \ldots, M$, while the off-diagonal entries represent the covariance between two different variables as shown in Equation (1).
- A positive value in covariance matrix means a positive correlation between the two variables, while the negative value indicates a negative correlation and zero value indicate that the two variables are uncorrelated or statistically independent.

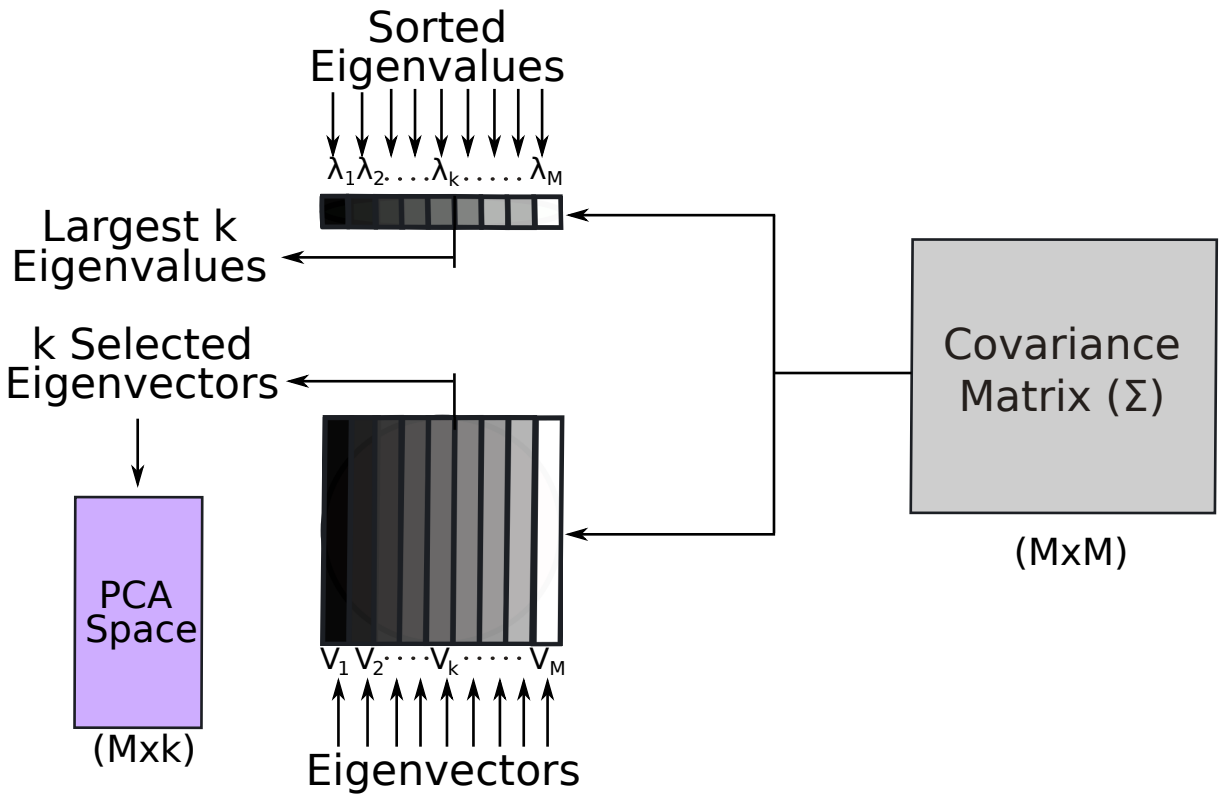
$$\begin{pmatrix} Var(x_1, x_1) & Cov(x_1, x_2) & \ldots & Cov(x_1, x_M) \\ Cov(x_2, x_1) & Var(x_2, x_2) & \ldots & Cov(x_2, x_M) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(x_M, x_1) & Cov(x_M, x_2) & & Var(x_M, x_M) \end{pmatrix} \quad (1)$$

[3] X is positive semi-definite if $v^T X v \geq 0$ for all $v \neq 0$. In other words, all eigenvalues of X are ≥ 0.

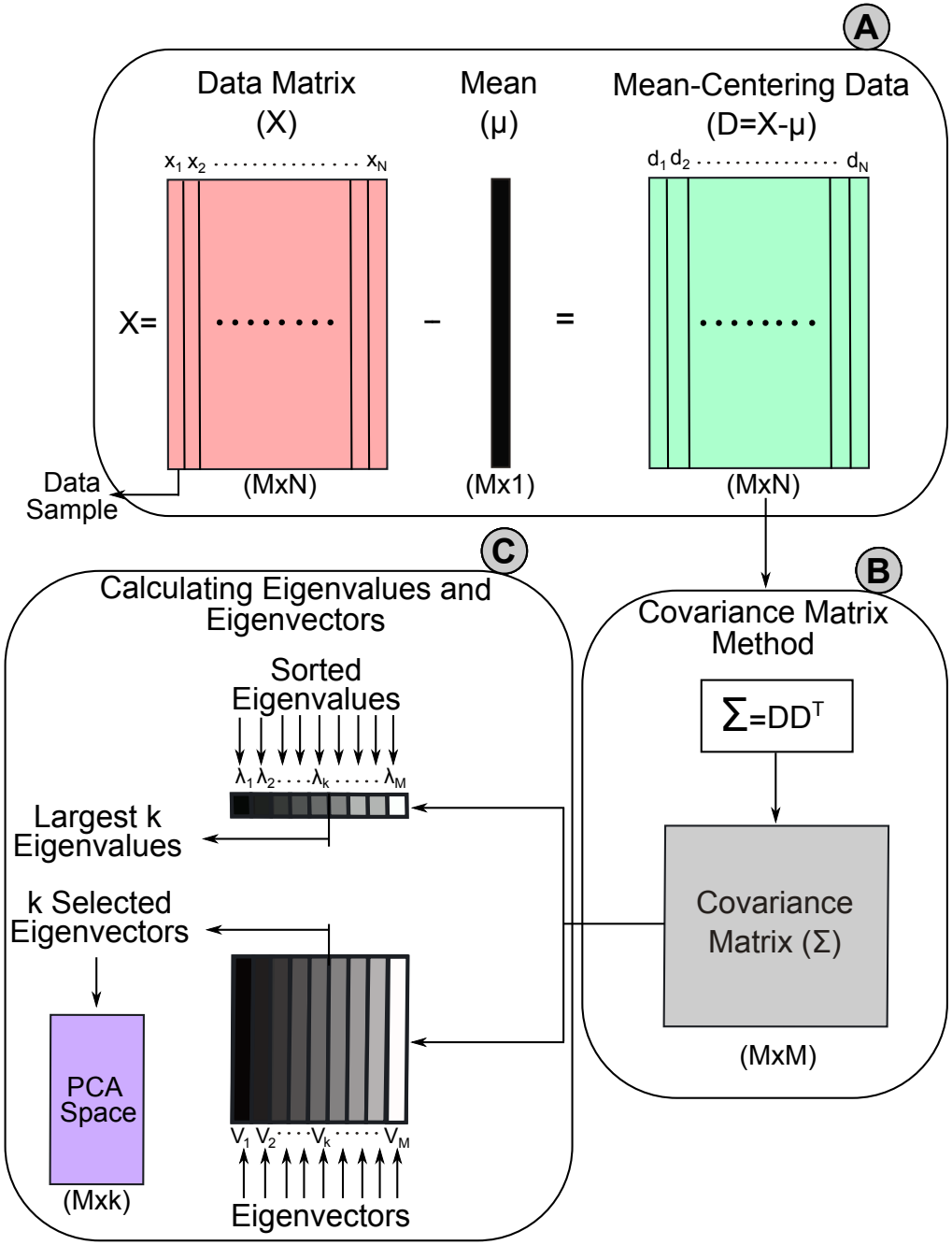
- The covariance matrix is solved by calculating the eigenvalues (λ) and eigenvectors (V) as follows:

$$V\Sigma = \lambda V \tag{2}$$

  - where V and λ represent the eigenvectors and eigenvalues of the covariance matrix, respectively.



- Eigenvalues are scalar values, while the eigenvectors are non-zero vectors, which represent the principal components, i.e. each eigenvector represents one principal component.
- Eigenvectors represent the directions of the PCA space, and the corresponding eigenvalues represent the scaling factor, length, magnitude, or the robustness of the eigenvectors.
- The eigenvector with the highest eigenvalue represents the first principal component and it has the maximum variance.
- The eigenvalues may be equal when the PCs have equal variances and hence all the eigenvectors are the same and we cannot decide which eigenvectors are used to construct the PCA space.


A
Data Matrix
(X)
Mean
(μ)
Mean-Centering Data
(D=X-μ)
x1 x2 ..... xN
d1 d2 ..... dN
X=
–
=
Data
Sample
(MxN)
(Mx1)
(MxN)
B
Covariance Matrix
Method
Σ=DDT
Covariance
Matrix (Σ)
(MxM)
C
Calculating Eigenvalues and
Eigenvectors
Sorted
Eigenvalues
λ1 λ2 ... λk ..... λM
Largest k
Eigenvalues
k Selected
Eigenvectors
PCA
Space
(Mxk)
V1 V2 ... Vk ..... VM
Eigenvectors

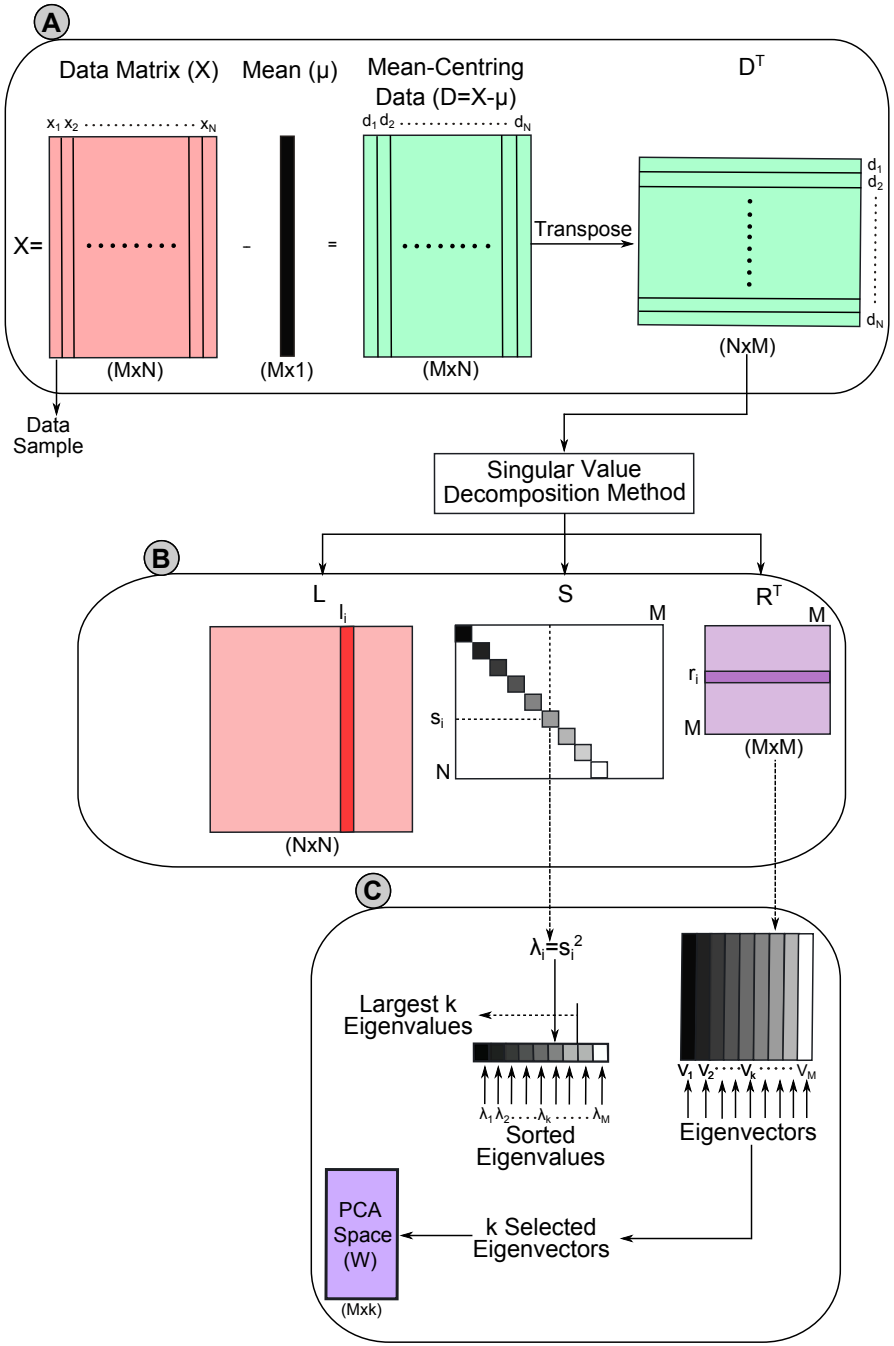
- Introduction to Principal component analysis (PCA)
  - Principal Components (PCs)
  - Covariance Matrix Method
  - Singular Value Decomposition (SVD) Method
  - PCA Space (Lower Dimensional Space)
  - Data Reconstruction
  - PCA Algorithms
- Numerical example
  - First Example: 2D-Class Example
  - Multi-Class Example
- Experimental Results and Discussion
  - Biometric Experiment
  - Image Compression Experiment
  - Data Visualization Experiment
- Conclusions

- Singular value decomposition is one of the most important linear algebra principles. The aim of the SVD method is to diagonalize the data matrix ($X \in \mathcal{R}^{p \times q}$) into three matrices.

$$X = LSR^T = \begin{bmatrix} l_1 & \cdots & l_M \end{bmatrix} \begin{bmatrix} s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & s_N \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -r_1^T- \\ -r_2^T- \\ \vdots \\ -r_N^T- \end{bmatrix} \quad (3)$$

  - $L(p \times p)$ are called left singular vectors,
  - $S(p \times q)$ is a diagonal matrix represents the singular values that are sorted from high-to-low, i.e. the highest singular value in the upper-left index of S, thus, $s_1 \geq s_2 \geq \cdots \geq s_q \geq 0$,
  - $R(q \times q)$ represents the right singular vectors.

- The left and right singular matrices, i.e. L and R, are orthonormal bases. To calculate SVD, R^T and S are first calculated by diagonalizing X^TX as follows,
  $X^TX = (LSR^T)^T(LSR^T) = RS^TL^TLSR^T = RS^2R^T$, where $L^TL = I$.
- The left singular vectors (L) is then calculated as follows,
  $L = XRS^{-1}$, where Xr_i is in the direction of s_il_i.
- The columns of the right singular vectors (R) represent the eigenvectors of X^TX or the principal components of the PCA space, and s_i^2, $\forall\, i = 1, 2, \ldots, q$ represent their corresponding eigenvalues.
- Since, the number of principal components and their eigenvalues are equal to q, thus the dimension of our original data matrix must be reversed to be compatible with SVD method. In other words, the mean-centering matrix is transposed before calculating the SVD method and hence each sample is represented by one row.


A
Data Matrix (X)
Mean (μ)
Mean-Centring Data (D=X-μ)
D^T
X=
(MxN)
(Mx1)
(MxN)
Transpose
(NxM)
Data Sample
Singular Value Decomposition Method
B
L
S
R^T
(NxN)
(MxM)
C
λi=si^2
Largest k Eigenvalues
Sorted Eigenvalues
Eigenvectors
PCA Space (W)
k Selected Eigenvectors
(Mxk)

- To compute the PCA space, eigenvalues and eigenvectors of the covariance matrix are calculated, where the covariance matrix is the product of DD^T, where $D = \{d_i\}_{i=1}^{N}, d_i = x_i - \mu$.
- Calculating the SVD for the covariance matrix is as follows:

$$DD^T = (LSR^T)^T(LSR^T) = RS^TL^TLSR^T$$

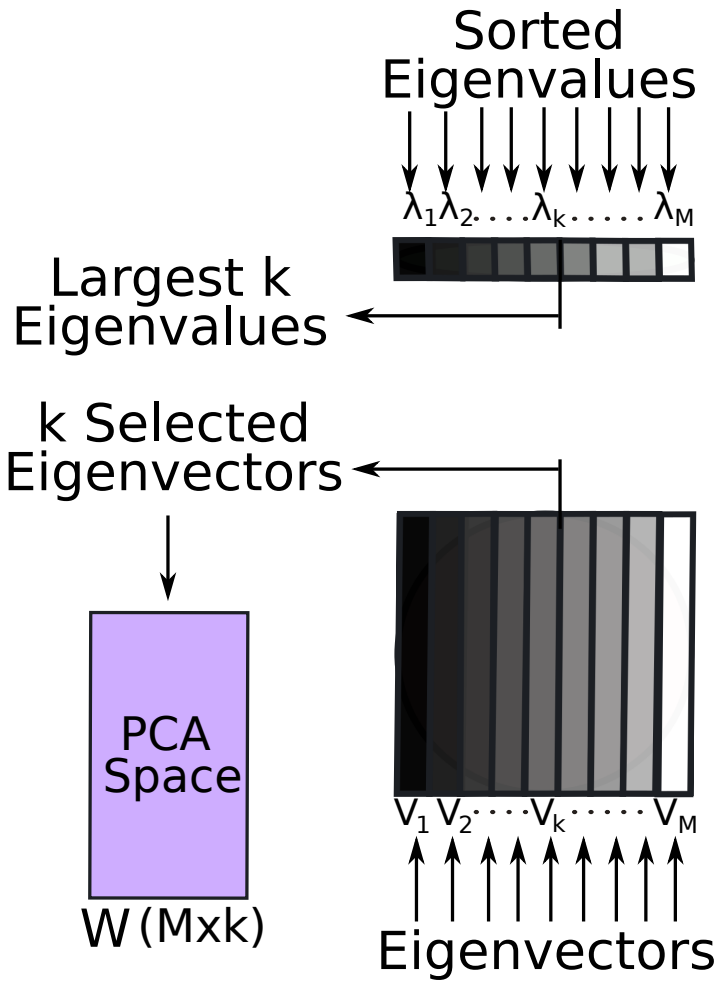
  - where $L^TL = I$

$$DD^T = RS^2R^T = (SVD(D^T))^2$$

    - where S^2 represents the eigenvalues of D^TD or DD^T and the columns of the right singular vector (R) represent the eigenvectors of DD^T.
- Hence, the square root of the eigenvalues that are calculated using the covariance matrix method are equal to the singular values of SVD method. Moreover, the eigenvectors of Σ are equal to the columns of R. Thus, the eigenvalues and eigenvectors that are calculated using the two methods are equal.

- Introduction to Principal component analysis (PCA)
  - Principal Components (PCs)
  - Covariance Matrix Method
  - Singular Value Decomposition (SVD) Method
  - PCA Space (Lower Dimensional Space)
  - Data Reconstruction
  - PCA Algorithms
- Numerical example
  - First Example: 2D-Class Example
  - Multi-Class Example
- Experimental Results and Discussion
  - Biometric Experiment
  - Image Compression Experiment
  - Data Visualization Experiment
- Conclusions

- To construct the lower dimensional space of PCA (W), a linear combination of k selected PCs that have the most k eigenvalues are used to preserve the maximum amount of variance, i.e. preserve the original data, while the other eigenvectors or PCs are neglected

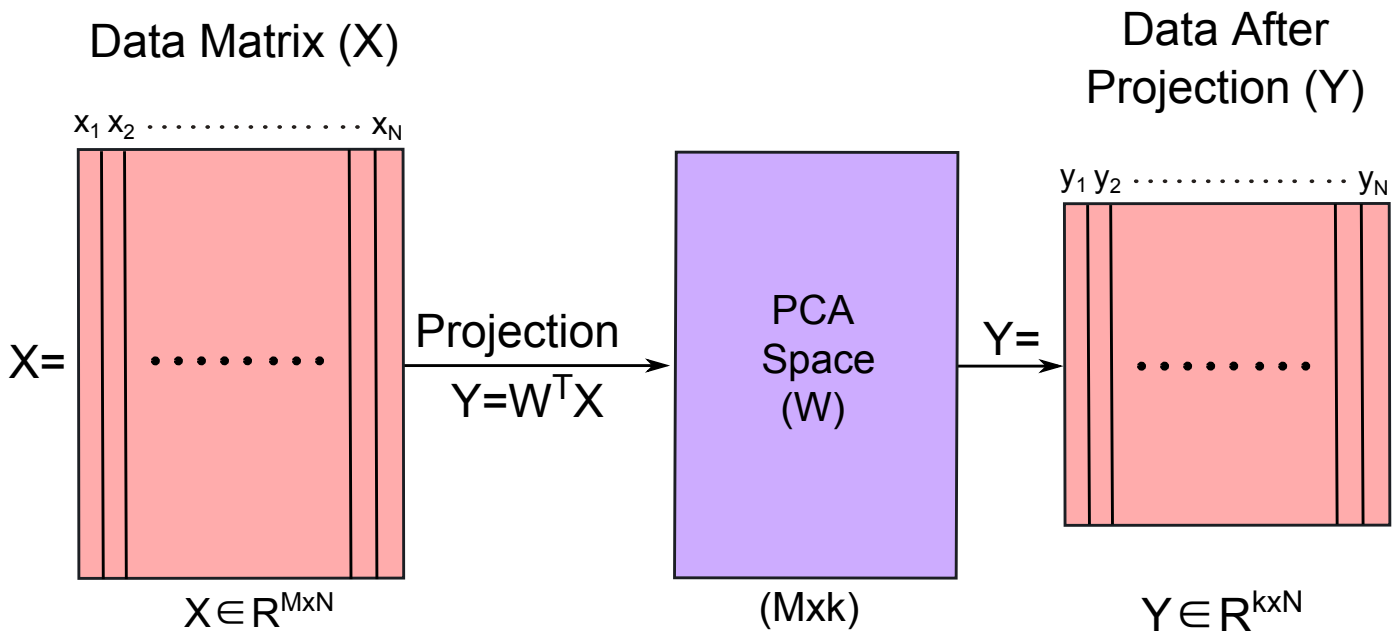


- The lower dimensional space is denoted by $W = \{v_1, \ldots, v_k\}$. The dimension of the original data is reduced by projecting it after subtracting the mean onto the PCA space as follows:

$$Y = W^T D = \sum_{i=1}^{N} W^T (x_i - \mu) \tag{4}$$

  - where $Y \in \mathcal{R}^k$ represents the original data after projecting it onto the PCA space; thus; $(M - k)$ features or variables are lost from the original data.

- The original data are projected on the PCA space, and the PCA reduces the dimension from M to k.



- Introduction to Principal component analysis (PCA)
  - Principal Components (PCs)
  - Covariance Matrix Method
  - Singular Value Decomposition (SVD) Method
  - PCA Space (Lower Dimensional Space)
  - Data Reconstruction
  - PCA Algorithms
- Numerical example
  - First Example: 2D-Class Example
  - Multi-Class Example
- Experimental Results and Discussion
  - Biometric Experiment
  - Image Compression Experiment
  - Data Visualization Experiment
- Conclusions

- The original data can be reconstructed again as follows:

$$\hat{X} = WY + \mu = \sum_{i=1}^{N} W y_i + \mu \tag{5}$$

  - where $\hat{X}$ represents the reconstructed data.
- The deviation between the original data and the reconstructed data are called the reconstruction error or residuals as follows:

$$Error = X - \hat{X} = \sum_{i=1}^{N} (x_i - \hat{x_i})^2 \tag{6}$$

$$Error = X - \hat{X} = \sum_{i=1}^{N} (x_i - \hat{x}_i)^2 \tag{7}$$

- The reconstruction error represents the square distance between the original data and the reconstructed data, and it is inversely proportional to the total variance of the PCA space. Thus, selecting a large number of PCs, increases the total variance of W and decreases the error between the reconstructed and the original data.
  - $\Rightarrow$ the robustness of the PCA is controlled by the number of selected eigenvectors (k) and it is measured by the sum of the selected eigenvalues, which is called total variance.
- The robustness of the lower dimensional space $W = \{v_1 \ldots, v_k\}$ is measured by the ratio between the total variance $(\lambda_i, i = 1, \ldots, k)$ of W to the total variance.

$$\text{Robustness of the PCA space} = \frac{\text{Total Variance of } W}{\text{Total Variance}} = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{M} \lambda_i} \tag{8}$$

- Introduction to Principal component analysis (PCA)
  - Principal Components (PCs)
  - Covariance Matrix Method
  - Singular Value Decomposition (SVD) Method
  - PCA Space (Lower Dimensional Space)
  - Data Reconstruction
  - PCA Algorithms
- Numerical example
  - First Example: 2D-Class Example
  - Multi-Class Example
- Experimental Results and Discussion
  - Biometric Experiment
  - Image Compression Experiment
  - Data Visualization Experiment
- Conclusions

Algorithm 1 : Calculating PCs using Covariance Matrix Method.

1: Given a data matrix ($X = [x_1, x_2, \ldots, x_N]$), where N represents the total number of samples and x_i represents the i^{th} sample.
2: Compute the mean of all samples as follows, $\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$.
3: Subtract the mean from all samples as follows, $D = \{d_1, d_2, \ldots, d_N\} = \sum_{i=1}^{N} x_i - \mu$.
4: Compute the covariance matrix as follows, $\Sigma = \frac{1}{N-1} D \times D^T$.
5: Compute the eigenvectors V and eigenvalues λ of the covariance matrix (Σ).
6: Sort eigenvectors according to their corresponding eigenvalues.
7: Select the eigenvectors that have the largest eigenvalues $W = \{v_1, \ldots, v_k\}$. The selected eigenvectors (W) represent the projection space of PCA.
8: All samples are projected on the lower dimensional space of PCA (W) as follows, $Y = W^T D$.

Algorithm 2 : Calculating PCs using SVD Method.

1: Given a data matrix ($X = [x_1, x_2, \ldots, x_N]$), where N represents the total number of samples and $x_i(M \times 1)$ represents the i^{th} sample.
2: Compute the mean of all samples as follows, $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$.
3: Subtract the mean from all samples as follows, $D = \{d_1, d_2, \ldots, d_N\} = \sum_{i=1}^{N} x_i - \mu$.
4: Construct a matrix $Z = \frac{1}{\sqrt{N-1}} D^T$, $Z(N \times M)$.
5: Calculate SVD for Z matrix as in Equation (3).
6: The diagonal elements of S represent the square root of the sorted eigenvalues, $\lambda = diag(S^2)$, while the PCs are represented by the columns of R.
7: Select the eigenvectors that have the largest eigenvalues $W = \{R_1, R_2, \ldots, R_k\}$ to construct the PCA space.
8: All samples are projected on the lower dimensional space of PCA (W) as follows, $Y = W^T D$.

- Introduction to Principal component analysis (PCA)
  - Principal Components (PCs)
  - Covariance Matrix Method
  - Singular Value Decomposition (SVD) Method
  - PCA Space (Lower Dimensional Space)
  - Data Reconstruction
  - PCA Algorithms
- Numerical example
  - First Example: 2D-Class Example
  - Multi-Class Example
- Experimental Results and Discussion
  - Biometric Experiment
  - Image Compression Experiment
  - Data Visualization Experiment
- Conclusions

- In this section, two numerical examples were illustrated to calculate the lower dimensional space.
- In the first example, the samples were represented by only two features to visualize it and the PCs were calculated using the two methods, i.e. covariance matrix and SVD methods.
- In the second example, each sample was represented by four features to show how the steps of PCA were affected by changing the dimension. Moreover, in this example, the influences of a constant variable, i.e. zero variance, were explained. MATLAB codes for all experiments are introduced in Appendix A.1 (in the paper).

- Introduction to Principal component analysis (PCA)
  - Principal Components (PCs)
  - Covariance Matrix Method
  - Singular Value Decomposition (SVD) Method
  - PCA Space (Lower Dimensional Space)
  - Data Reconstruction
  - PCA Algorithms
- Numerical example
  - First Example: 2D-Class Example
  - Multi-Class Example
- Experimental Results and Discussion
  - Biometric Experiment
  - Image Compression Experiment
  - Data Visualization Experiment
- Conclusions

- The goal of this experiment is to calculate principal components using covariance matrix method
- Given a data matrix $X = \{x_1, x_2, \ldots, x_8\}$, where x_i represents the i^{th} samples as denoted in Equation (9). Each sample of the matrix was represented by one column that consists of two features ($x_i \in \mathcal{R}^2$) to visualize it. The total mean (μ) was then calculated and its value was $\mu = \begin{bmatrix} 2.63 \\ 3.63 \end{bmatrix}$.

$$X = \begin{bmatrix} 1.00 & 1.00 & 2.00 & 0.00 & 5.00 & 4.00 & 5.00 & 3.00 \\ 3.00 & 2.00 & 3.00 & 3.00 & 4.00 & 5.00 & 5.00 & 4.00 \end{bmatrix} \quad (9)$$

- The data were then subtracted from the mean and the values of D will be as follows:

$$D = \begin{bmatrix} -1.63 & -1.63 & -0.63 & -2.63 & 2.38 & 1.38 & 2.38 & 0.38 \\ -0.63 & -1.63 & -0.63 & -0.63 & 0.38 & 1.38 & 1.38 & 0.38 \end{bmatrix}$$

- The covariance matrix (Σ) were then calculated. The eigenvalues (λ) and eigenvectors (V) of the covariance matrix were then calculated. The values of the Σ, λ, and V are shown below.

$$\Sigma = \begin{bmatrix} 3.70 & 1.70 \\ 1.70 & 1.13 \end{bmatrix}, \lambda = \begin{bmatrix} 0.28 & 0.00 \\ 0.00 & 4.54 \end{bmatrix}, \text{ and } V = \begin{bmatrix} 0.45 & -0.90 \\ -0.90 & -0.45 \end{bmatrix}$$

- From the results, we note that the second eigenvalue (λ_2) was more than the first one (λ_1).
- The second eigenvalue represents $\frac{4.54}{0.28+4.54} \approx 94.19\%$ of the total eigenvalues, i.e. total variance, while the first eigenvalue represents $\frac{0.28}{0.28+4.54} \approx 5.81\%$, which reflects the robustness of the second eigenvector than the first one. Thus, the second eigenvector (i.e. second column of V) points to the direction of the maximum variance and hence it represents the first principal component of the PCA space.

- To calculating the projected data, first the mean-centering data (i.e. data − total mean) were then projected on each eigenvector as follows, ($Y_{v1} = v_1^T D$ and $Y_{v2} = v_2^T D$), where Y_{v1} and Y_{v2} represent the projection of the D on the first and second eigenvectors, i.e. v_1 and v_2, respectively.
- The values of Y_{v1} and Y_{v1} are shown below.

$$Y_{v1} = \begin{bmatrix} -0.16 & 0.73 & 0.28 & -0.61 & 0.72 & -0.62 & -0.18 & -0.17 \end{bmatrix}$$

$$Y_{v2} = \begin{bmatrix} 1.73 & 2.18 & 0.84 & 2.63 & -2.29 & -1.84 & -2.74 & -0.50 \end{bmatrix}$$

- The reconstruction error between the original data and the reconstructed data using all eigenvectors or all principal components approximately tend to zero, i.e. there is no information lost. In other words, if the original data are projected on all eigenvectors without neglecting anyone, and then reconstructed again, the error between the original and the reconstructed data will be zero.
- However, removing one or more eigenvectors to construct a lower dimensional space ($W \in \mathcal{R}^k$), reduces the dimension of the original data to k, thus some data are neglected, i.e. when the dimension of the lower dimensional space is lower than the original dimension, hence, there is a difference between the original data and the reconstructed data.
- The reconstruction error or residual depends on the number of the selected eigenvectors (k) and the robustness of those eigenvectors, which is measured by their corresponding eigenvalues.

- In this example, the reconstruction error was calculated when the original data were first reconstructed as in Equation (5).
- The original data were projected on the two calculated eigenvectors separately, i.e. $Y_{v1} = v_1^T D$ and $Y_{v2} = v_2^T D$.
- Each eigenvector represents a separate lower dimensional space.
- The original data were then reconstructed using the same eigenvector that was used in the projection as follows, $\hat{X}_i = v_i Y_{vi} + \mu$.
- The values of the reconstructed data (i.e. $\hat{X}_1$ and $\hat{X}_2$) are as follows:

$$\hat{X}_1 = v_1 Y_{v1} + \mu = \begin{bmatrix} 2.55 & 2.95 & 2.75 & 2.35 & 2.95 & 2.35 & 2.55 & 2.55 \\ 3.77 & 2.97 & 3.37 & 4.17 & 2.98 & 4.18 & 3.78 & 3.78 \end{bmatrix}$$

$$\hat{X}_2 = v_2 Y_{v2} + \mu = \begin{bmatrix} 1.07 & 0.67 & 1.88 & 0.27 & 4.68 & 4.28 & 5.08 & 3.08 \\ 2.85 & 2.66 & 3.25 & 2.46 & 4.65 & 4.45 & 4.84 & 3.85 \end{bmatrix}$$

- The error between the original data and the reconstructed data that were projected on the first and second eigenvectors (E_{v1} and E_{v2}) are

$$E_{v1} = X - \hat{X}_1 = \begin{bmatrix} -1.55 & -1.95 & -0.75 & -2.35 & 2.05 & 1.65 & 2.45 & 0.45 \\ -0.77 & -0.97 & -0.37 & -1.17 & 1.02 & 0.82 & 1.22 & 0.22 \end{bmatrix}$$

$$E_{v2} = X - \hat{X}_2 = \begin{bmatrix} -0.07 & 0.33 & 0.12 & -0.27 & 0.32 & -0.28 & -0.08 & -0.08 \\ 0.15 & -0.66 & -0.25 & 0.54 & -0.65 & 0.55 & 0.16 & 0.15 \end{bmatrix}$$

- The error between the original data and the reconstructed data that were projected on the eigenvector (E_{v2}) was much lower than the reconstructed data that were projected on the first eigenvector, E_{v1}.
- The total error between the reconstructed data that were projected on the first eigenvector and the original data was $3.00 + 4.75 + 0.7 + 6.91 + 5.26 + 3.4 + 7.5 + 0.25 \approx 31.77$, while the error using the second eigenvector was equal to $0.03 + 0.54 + 0.08 + 0.37 + 0.52 + 0.38 + 0.03 + 0.03 \approx 1.98$[4].

[4] For example, the error between the first sample ($[1\ \ 3]^T$) and the reconstructed sample using v_1 and v_2 was, $(2.55-1)^2 + (3.77-3)^2 \approx 3.00$ and $(1.07-1)^2 + (2.85-3)^2 \approx 0.03$, respectively.

- This figure shows the error E_{v1} and E_{v2} in green and blue lines. As shown, E_{v2} was much lower than E_{v1}.

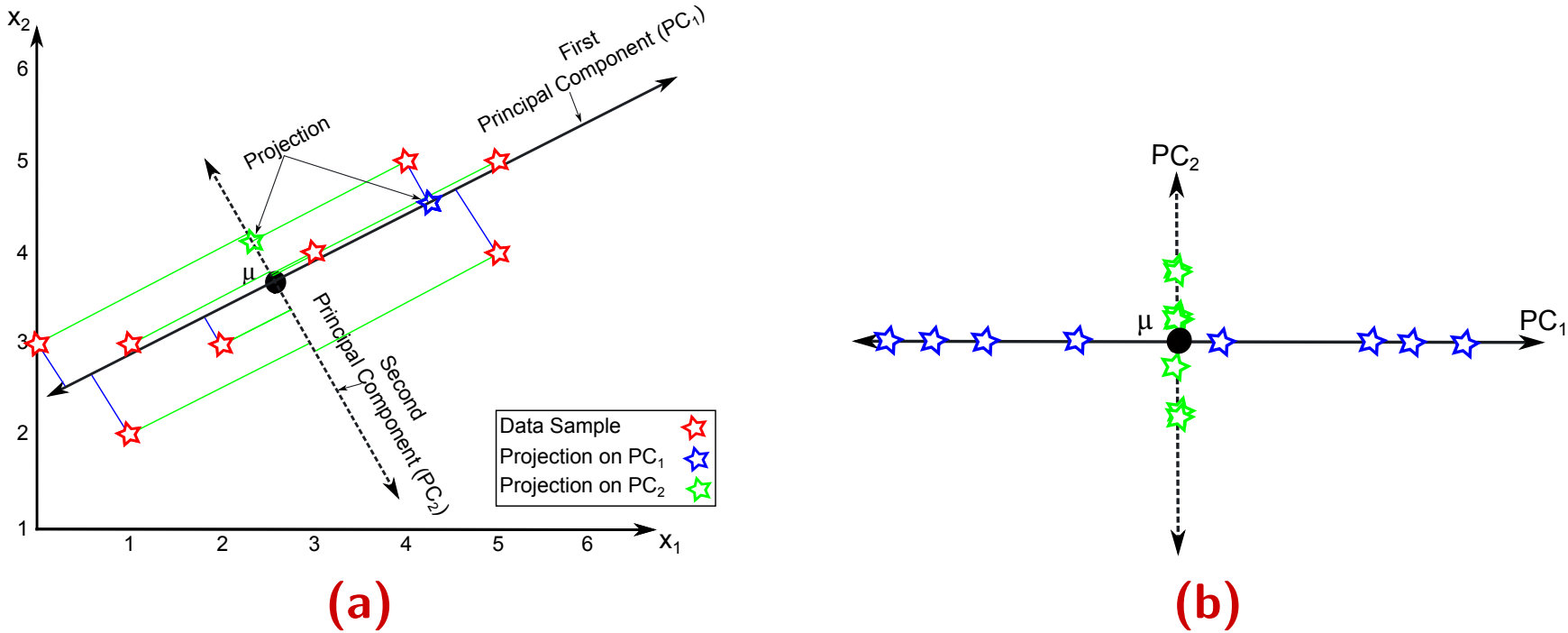


Figure: A visualized example of the PCA technique, (a) the dotted line represents the first eigenvector (v_1), while the solid line represents the second eigenvector (v_2) and the blue and green lines represent the reconstruction error using PC_1 and PC_2, respectively; (b) projection of the data on the principal components, the blue and green stars represent the projection onto the first and second principal components, respectively.

- We can also calculate the lower dimensional space of PCA using SVD method.
- The original data (X) that were used in the covariance matrix example were used in this example.
- The first three steps in SVD method and covariance matrix methods are common.
- In the fourth step in SVD, the original data were transposed as follows, $Z = \frac{1}{N-1} D^T$. The values of Z are as follows:

$$Z = \begin{bmatrix} -0.61 & -0.24 \\ -0.61 & -0.61 \\ -0.24 & -0.24 \\ -0.99 & -0.24 \\ 0.90 & 0.14 \\ 0.52 & 0.52 \\ 0.90 & 0.52 \\ 0.14 & 0.14 \end{bmatrix}$$

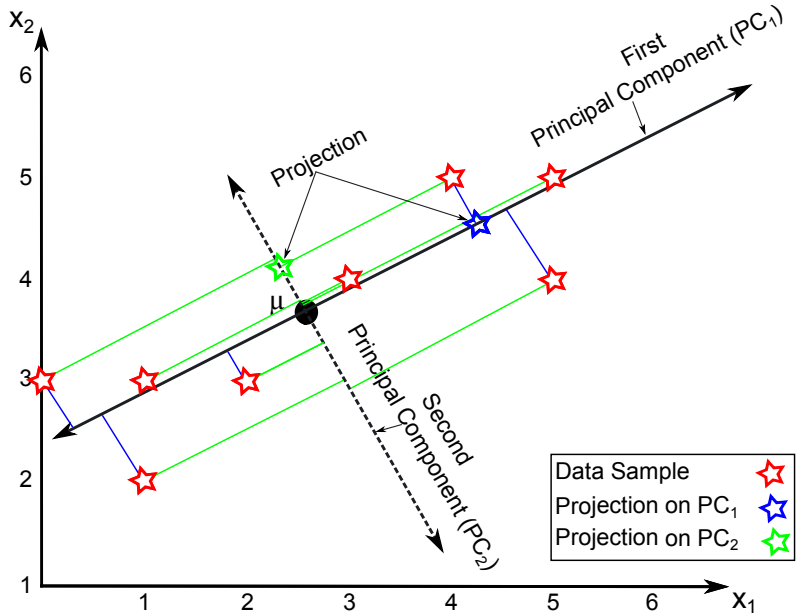
- SVD was then used to calculate L, S, and R as in Equation (3) and their values are as follows:

$$L = \begin{bmatrix} -0.31 & 0,12 & -0,07 & -0,60 & 0,58 & 0,15 & 0,41 & 0,04 \\ -0.39 & -0,52 & -0,24 & 0,20 & -0,29 & 0,53 & 0,31 & 0,14 \\ -0.15 & -0,20 & 0,96 & -0,01 & -0,01 & 0,08 & 0,07 & 0,02 \\ -0.47 & 0,43 & 0,02 & 0,69 & 0,32 & -0,05 & 0.12 & -0,01 \\ 0.41 & -0,51 & -0,04 & 0,31 & 0,68 & 0,08 & -0,09 & 0,02 \\ 0.33 & 0.44 & 0,08 & 0,02 & 0,02 & 0,82 & -0,15 & -0,05 \\ 0,49 & 0,12 & 0,05 & 0,17 & -0,15 & -0,12 & 0,83 & -0,03 \\ 0,09 & 0,12 & 0,02 & 0,00 & 0,01 & -0,05 & -0,04 & 0,99 \end{bmatrix}$$

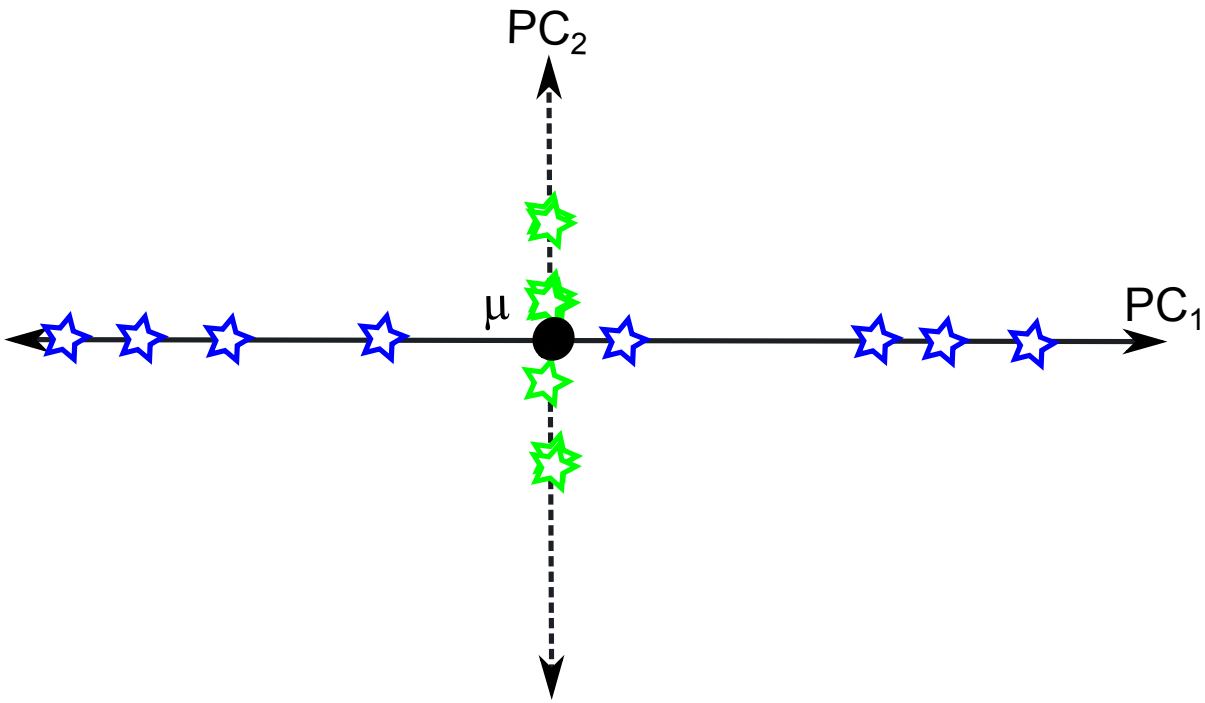
$$S = \begin{bmatrix} 2.13 & 0 \\ 0 & 0.53 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, R = \begin{bmatrix} 0.90 & -0.45 \\ 0.45 & 0.90 \end{bmatrix}$$

- As shown first singular value (s_1) was more than the second one, thus, the first column of R represents the first principal component, and it was in the same direction of the second column of V that was calculated using the covariance matrix method.
- The square roots of the eigenvalues that were calculated using the covariance matrix method were equal to the singular values of SVD. As a result, the eigenvalues and eigenvectors of the covariance matrix and SVD methods were the same.

- The figure below shows a comparison between the two principal components or eigenvectors.
- The samples of the original data are represented by red stars, each sample is represented by two features only ($x_i \in \mathcal{R}^2$) to be visualized.
- Hence, there are two eigenvectors (v_1 and v_2) are calculated using SVD or solving covariance matrix.
- The solid line represents the second eigenvector (v_2), i.e. first principal component, while the dotted line represents the first eigenvector (v_1), i.e. second principal component.

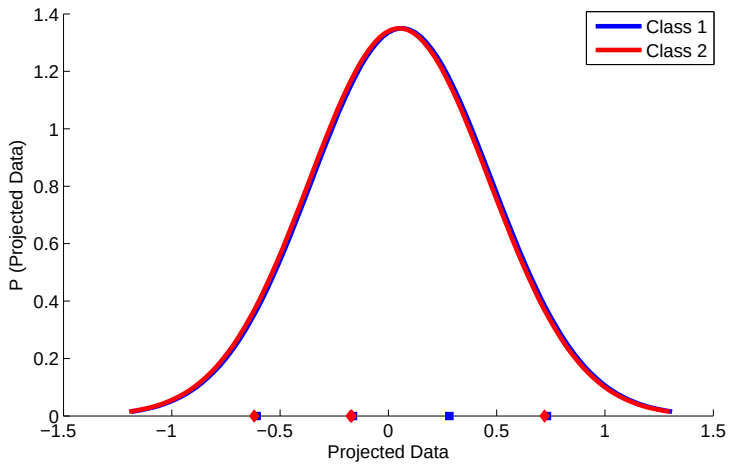


- The figure below shows the projection of the original data on the two principal components.

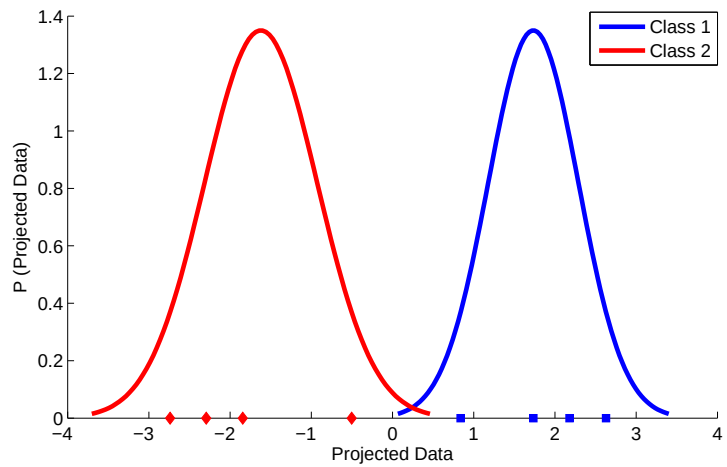


- PCA was used to remove the redundant data or noise that have a good impact on the classification problem.
- Hence, PCA was used commonly as a feature extraction method.
- In this experiment, a comparison between the two eigenvectors was performed to show which one was suitable to construct a sub-space to discriminate between different classes. Hence, PCA can be usedas preprocessing step for classification problems.
- Assume that, the original data consists two classes. The first class ($\omega_1 = \{x_1, x_2, x_3, x_4\}$) consists of the first four observations or samples, while the other four samples represent the second class ($\omega_2 = \{x_5, x_6, x_7, x_8\}$).

- As shown in the below figure **(a)**, more important data or information that were used to discriminate between the two classes were lost and discarded when the data were projected on the second principal component.
- The important information of the original data were preserved when the data were projected on the first principal component, hence, the two classes can be discriminated as shown in **(b)**.
- Thus, the robust eigenvectors were preserved the important information that were used to classify between different classes.



(a)



(b)

- Introduction to Principal component analysis (PCA)
  - Principal Components (PCs)
  - Covariance Matrix Method
  - Singular Value Decomposition (SVD) Method
  - PCA Space (Lower Dimensional Space)
  - Data Reconstruction
  - PCA Algorithms
- Numerical example
  - First Example: 2D-Class Example
  - Multi-Class Example
- Experimental Results and Discussion
  - Biometric Experiment
  - Image Compression Experiment
  - Data Visualization Experiment
- Conclusions

- In this example, each sample was represented by four variables. In this experiment, the second variable was constant for all observations as shown below.

$$X = \begin{bmatrix} 1.00 & 1.00 & 2.00 & 0.00 & 7.00 & 6.00 & 7.00 & 8.00 \\ 2.00 & 2.00 & 2.00 & 2.00 & 2.00 & 2.00 & 2.00 & 2.00 \\ 5.00 & 6.00 & 5.00 & 9.00 & 1.00 & 2.00 & 1.00 & 4.00 \\ 3.00 & 2.00 & 3.00 & 3.00 & 4.00 & 5.00 & 5.00 & 4.00 \end{bmatrix}$$

- The covariance matrix of the given data was calculated and its values are shown below.

$$\Sigma = \begin{bmatrix} 10.86 & 0 & -7.57 & 2.86 \\ 0 & 0 & 0 & 0 \\ -7.57 & 0 & 7.55 & -2.23 \\ 2.86 & 0 & -2.23 & 1.13 \end{bmatrix}$$

- The values of second row and column were zeros, which reflects that the variance of the second variable was zeros because the second variable was constant.

- The eigenvalues and eigenvectors of the covariance matrix of the above data are shown below.

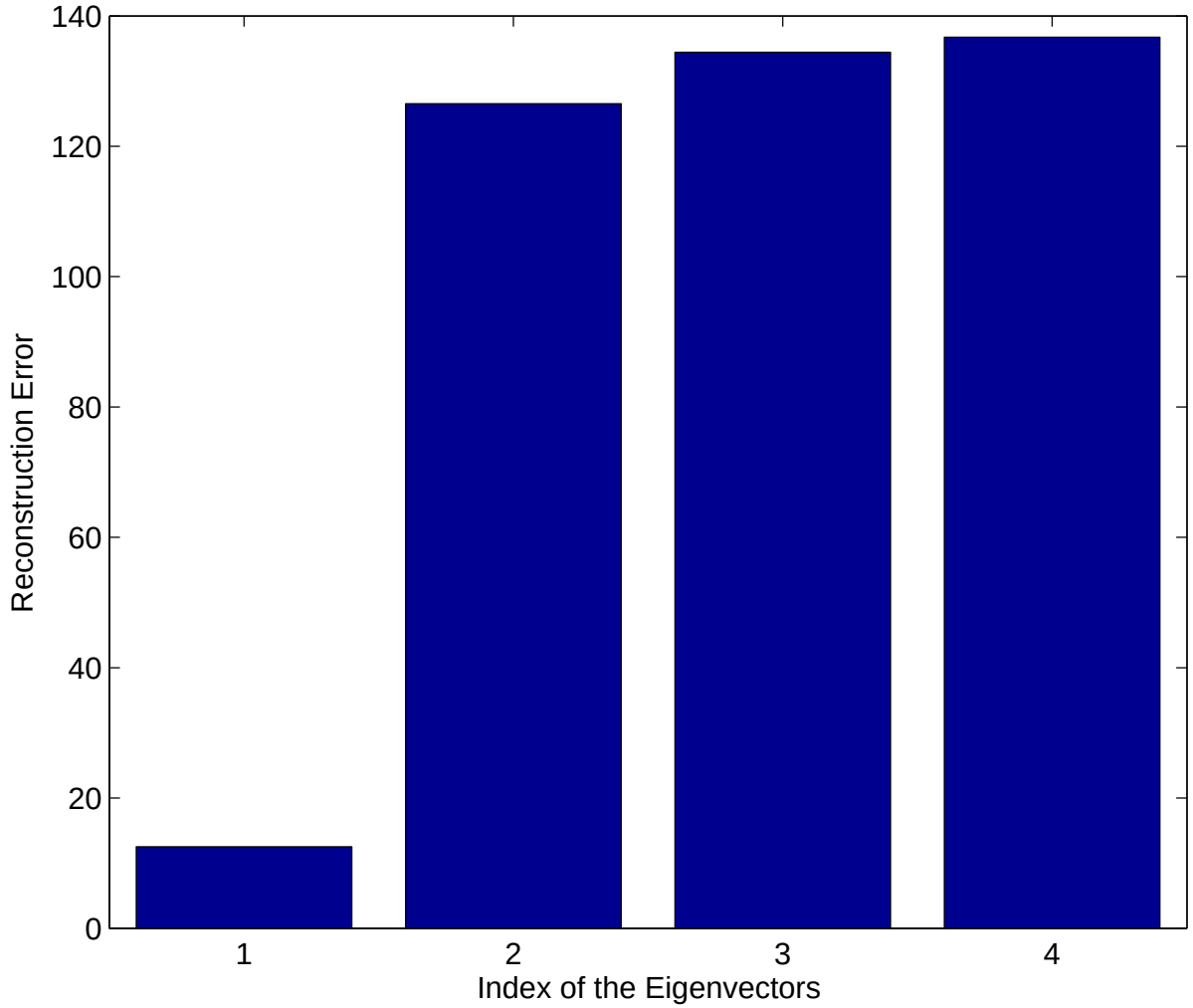
$$\lambda = \begin{bmatrix} 17.75 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.46 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.33 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix} \quad V = \begin{bmatrix} 0.76 & 0.62 & -0.20 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.00 \\ -0.61 & 0.79 & 0.10 & 0.00 \\ 0.21 & 0.05 & 0.98 & 0.00 \end{bmatrix}$$

- The first eigenvector represents the first principal component because it was equal to $\frac{17.75}{17.75+1.46+0.33+0} \approx 90.84\%$ of the total variance.
- The first three eigenvectors represent 100% of the total variance of the total data and the fourth eigenvector was redundant.
- The second variable, i.e. second row, will be neglected completely when the data were projected on any of the best three eigenvectors.

- The projected data on the fourth eigenvector preserved only the second variable and all the other original data were lost, and the reconstruction error was ≈ 136.75, while the reconstruction error was ≈ 12.53 , 126.54 , 134.43 when the data were projected on the first three eigenvectors, respectively.

$$Y_{v_1} = \begin{bmatrix} -2.95 & -3.78 & -2.19 & -6.16 & 4.28 & 3.12 & 4.49 & 3.20 \end{bmatrix}$$
$$Y_{v_2} = \begin{bmatrix} -1.20 & -0.46 & -0.58 & 1.32 & -0.58 & -0.39 & -0.54 & 2.39 \end{bmatrix}$$
$$Y_{v_3} = \begin{bmatrix} 0.06 & -0.82 & -0.14 & 0.64 & -0.52 & 0.75 & 0.45 & -0.43 \end{bmatrix}$$
$$Y_{v_4} = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix}$$

- The reconstruction error of the four eigenvectors in the multi-class example are shown below.



- Introduction to Principal component analysis (PCA)
  - Principal Components (PCs)
  - Covariance Matrix Method
  - Singular Value Decomposition (SVD) Method
  - PCA Space (Lower Dimensional Space)
  - Data Reconstruction
  - PCA Algorithms
- Numerical example
  - First Example: 2D-Class Example
  - Multi-Class Example
- Experimental Results and Discussion
  - Biometric Experiment
  - Image Compression Experiment
  - Data Visualization Experiment
- Conclusions

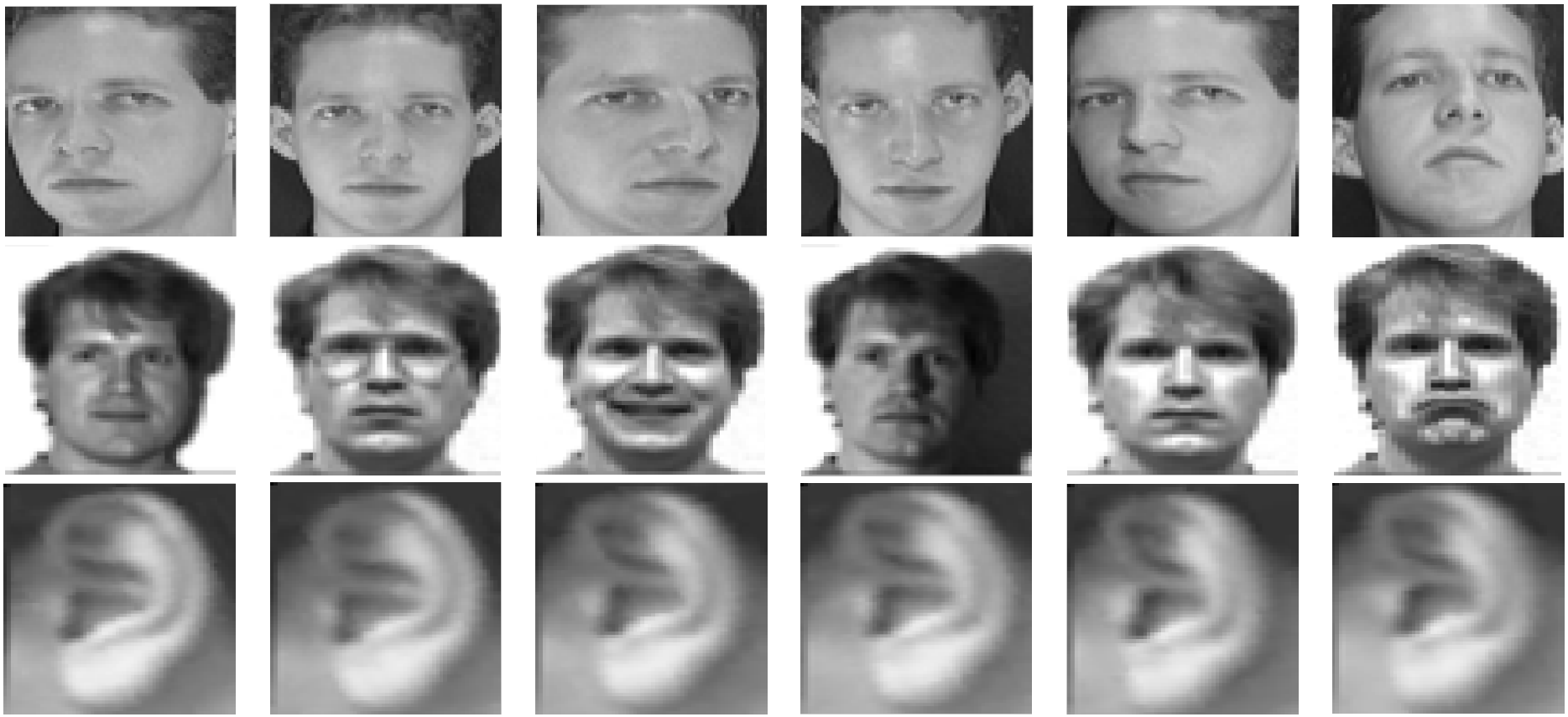
- In the first experiment, the PCA technique was used to identify individuals, i.e. PCA was used as a feature extraction method.
- In the second experiment, PCA was used to compress a digital image through removing the eigenvectors that represent the minimum variance from the PCA space and hence some redundant information will be removed from the original image.
- In this experiment, different numbers of eigenvectors were used to preserve the total information of the original image. In the third experiment, the PCA technique was used to reduce the dimension of the high-dimensional data to be visualized.

- The aim of this experiment was to investigate the impact of the number of principal components on the classification accuracy and the classification CPU time using these principal components. Three biometric datasets were used in this experiment. The descriptions of the datasets are as follows:
  - Olivetti Research Laboratory, Cambridge (ORL[5]) dataset, which consists of 40 individuals, each has ten grey scale images. The size of each image was 92×112.
  - Ear dataset images[6], which consists of 17 individuals, each has six grey scale images. The images have different dimensions, thus all images were resized to be 64×64.
  - Yale[7] face dataset images, which contains 165 grey scale images in GIF format of 15 individuals. Each individual has 11 images in different expressions and configuration: center-light, happy, left-light, with glasses, normal, right-light, sad, sleepy, surprised, and wink. The size of each image was 320×243.

[5] http://www.cam-orl.co.uk

[6] http://faculty.ucmerced.edu/mcarreira-perpinan/software.html

[7] http://vision.ucsd.edu/content/yale-face-database



- In this experiment, the Nearest Neighbor (NN) classifier was used. The nearest neighbor (minimum distance) classifier was used to classify the testing image by comparing its position in the PCA space with positions of the training images. The results of this experiment were evaluated using two different assessment methods, namely, accuracy and CPU time.

- In this experiment, different numbers of principal components were used to construct the PCA space. Hence, the dimension of the PCA space and the projected data were changed based on the number of selected principal components (k).

Table: A comparison between ORL, Ear, and Yale datasets in terms of accuracy (%), CPU time (sec), and cumulative variance (%) using different number of eigenvectors (biometric experiment).

Number of Eigenvectors	ORL Dataset			Ear Dataset			Yale Dataset		
	Acc. (%)	CPU Time (sec)	Cum. Var. (%)	Acc. (%)	CPU Time (sec)	Cum. Var. (%)	Acc. (%)	CPU Time (sec)	Cum. Var. (%)
1	13.33	0.074	18.88	15.69	0.027	29.06	26.67	0.045	33.93
5	80.83	0.097	50.17	80.39	0.026	66.10	76.00	0.043	72.24
10	94.17	0.115	62.79	90.20	0.024	83.90	81.33	0.042	85.13
15	95.00	0.148	69.16	94.12	0.028	91.89	81.33	0.039	90.18
20	95.83	0.165	73.55	94.12	0.033	91.89	84.00	0.042	93.36
30	95.83	0.231	79.15	94.12	0.033	98.55	85.33	0.061	96.60
40	95.83	0.288	82.99	94.12	0.046	99.60	85.33	0.064	98.22
50	95.83	0.345	85.75	94.12	0.047	100.00	85.33	0.065	99.12
100	95.83	0.814	93.08	94.12	0.061	100.00	85.33	0.091	100.00

- Introduction to Principal component analysis (PCA)
  - Principal Components (PCs)
  - Covariance Matrix Method
  - Singular Value Decomposition (SVD) Method
  - PCA Space (Lower Dimensional Space)
  - Data Reconstruction
  - PCA Algorithms
- Numerical example
  - First Example: 2D-Class Example
  - Multi-Class Example
- Experimental Results and Discussion
  - Biometric Experiment
  - Image Compression Experiment
  - Data Visualization Experiment
- Conclusions

- The aim of this experiment was to use the PCA technique as an image compression technique to reduce the size of the image by removing some redundant information.
- The image compression technique is divided into two techniques, namely, *lossy* and *lossless* compression techniques.
- In the lossless compression technique, all information that was originally in the original file remains after the file is uncompressed, i.e. all information is completely restored.
- In the lossy compression technique, some redundant information from the original file is lost. In this experiment, PCA was used to reduce the dimension of the original image and the original data was restored again..

- In this experiment, two images are used.

(a)

(b)

Figure: Original images (a) 512×512 8 bit/pixel original image of Lena, (b) 256×256 8 bit/pixel original image of Cameraman.

- Two different assessment methods were used to evaluate this experiment.
- The first assessment method was the Mean Square Error (MSE), which is the difference or error between the original image (I) and the reconstructed image ($\hat{I}$).

$$MSE = \frac{1}{rc}\sum_{i=1}^{r}\sum_{j=1}^{c}(I(i,j) - \hat{I}(i,j))^2$$

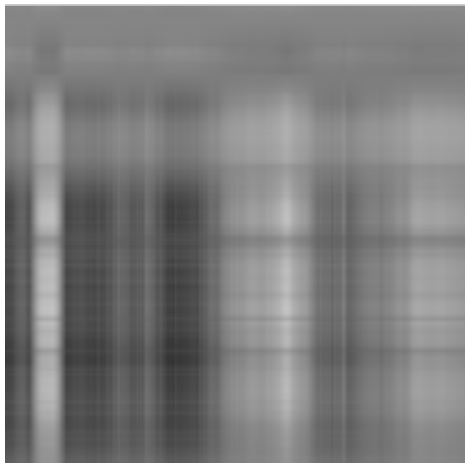
- The second assessment method was the compression ratio (CR) or the compression power, which is the ratio between the size of original image, i.e. the number of unit memory required to represent the original image, and the size of the compressed image

$$CR = \frac{\text{Uncompressed Size}}{\text{Compressed Size}}$$

- where I represents the original image, $\hat{I}$ is the reconstructed image, r and c represents the number of rows and columns of the image, respectively, i.e. r and c represent the dimension of the image.

Table: Compression ratio and mean square error of the compressed images using different percentages of the eigenvectors (image compression experiment).

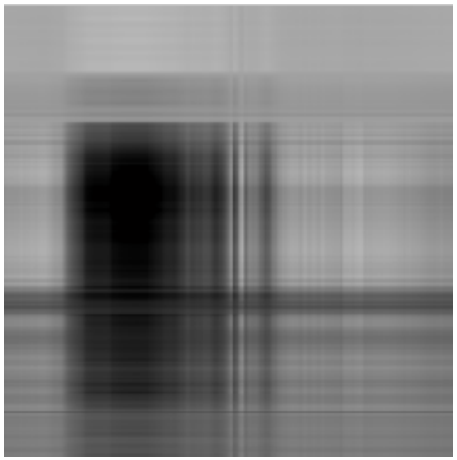
Percentage of the used Eigenvectors	Lena Image			Cameraman Image		
	MSE	CR	Cumulative Variance (%)	MSE	CR	Cumulative Variance (%)
10	5.3100	512:51.2	97.35	8.1057	256:25.6	94.56
20	2.9700	512:102.4	99.25	4.9550	256:51.2	98.14
30	1.8900	512:153.6	99.72	3.3324	256:76.8	99.24
40	1.3000	512:204.8	99.87	2.0781	256:102.4	99.73
50	0.9090	512:256	99.94	1.1926	256:128	99.91
60	0.6020	512:307.2	99.97	0.5588	256:153.6	99.98
70	0.3720	512:358.4	99.99	0.1814	256:179.2	100.00
80	0.1935	512:409.6	100.00	0.0445	256:204.8	100.00
90	0.0636	512:460.8	100.00	0.0096	256:230.4	100.00
100 (All)	0.0000	512:512=1	100.00	0.0000	1	100.00



(a) 1 PC, CR=512:1, MSE=30.7

(b) 10% of the PCs, CR=10:1, and MSE=5.31

(c) 50% of the PCs, CR=2:1, and MSE=0.909



(d) 1 PC, CR=256:1, and MSE=25.97

(e) 10% of the PCs, CR=10:1, and MSE=8.1057

(f) 50% of the PCs, CR=2:1, and MSE=1.1926

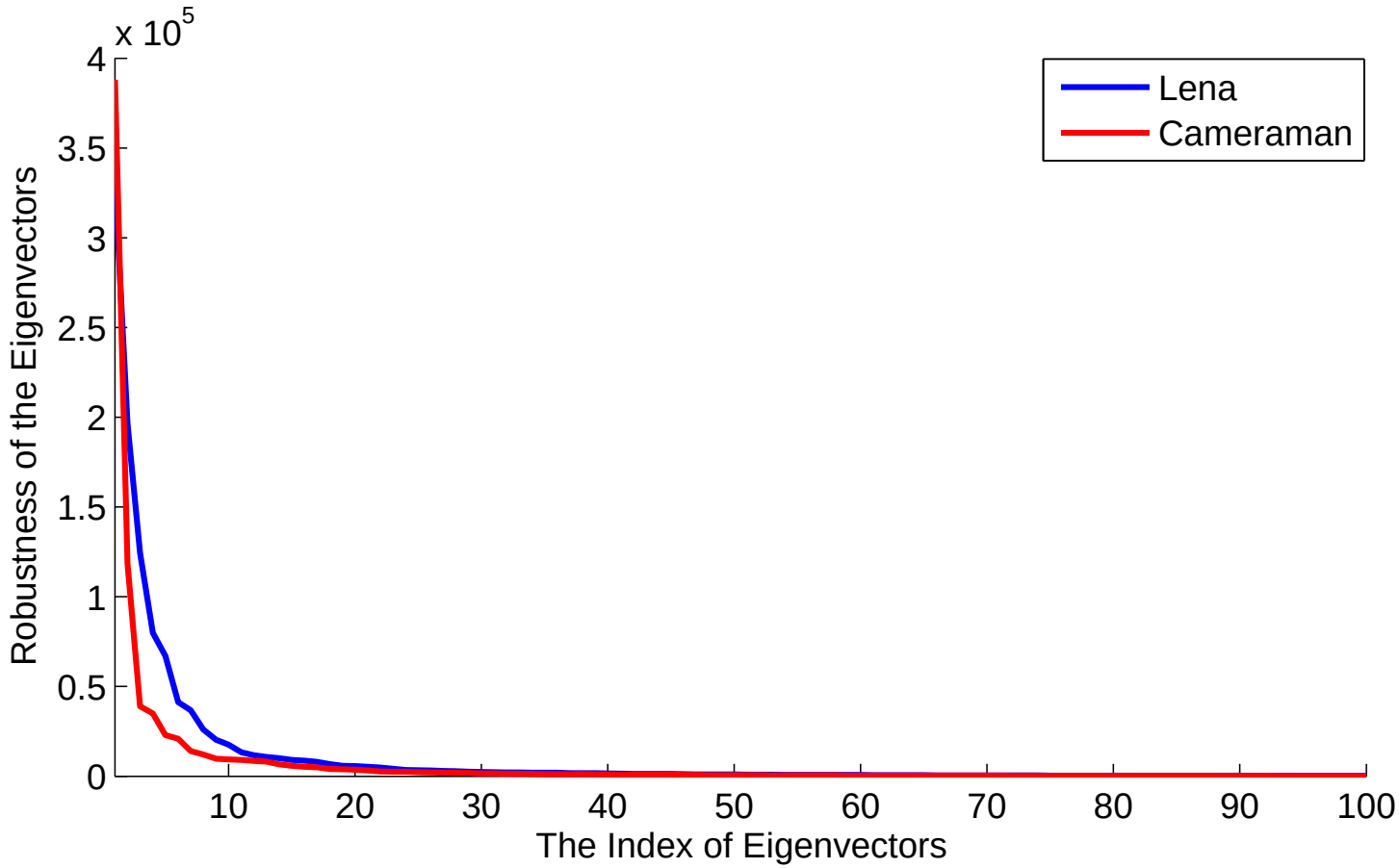


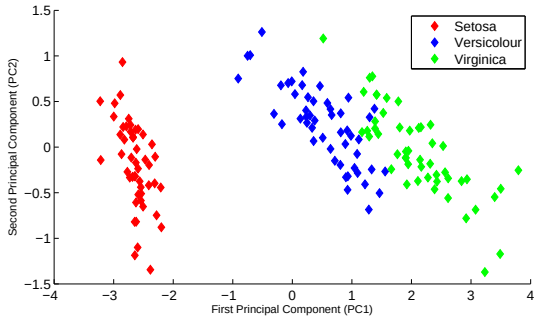
Figure: The robustness, i.e. total variance (see Equation (**??**)), of the first 100 eigenvectors using Lena and Cameraman images.

- Introduction to Principal component analysis (PCA)
  - Principal Components (PCs)
  - Covariance Matrix Method
  - Singular Value Decomposition (SVD) Method
  - PCA Space (Lower Dimensional Space)
  - Data Reconstruction
  - PCA Algorithms
- Numerical example
  - First Example: 2D-Class Example
  - Multi-Class Example
- Experimental Results and Discussion
  - Biometric Experiment
  - Image Compression Experiment
  - Data Visualization Experiment
- Conclusions

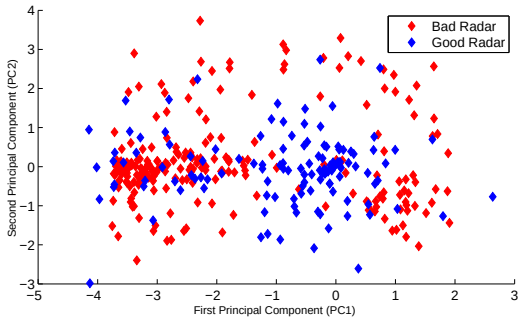
- The aim of this experiment was to use the PCA technique to reduce the dimension of the high dimensional datasets to be visualized.
- In this experiment, the PCA space was constructed using two or three principal components to visualize the datasets in 2D or 3D, respectively.
- Six standard datasets were used in this experiment, and each dataset has different numbers of attributes, classes, and samples.

Table: Datasets descriptions.

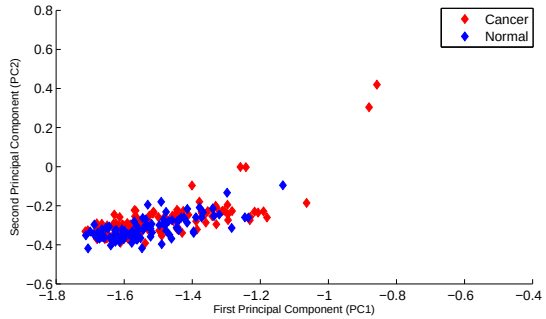
Dataset	Number of Classes	Number of Features	Number of Samples
Iris	3	4	150
Iono	2	34	351
Ovarian	2	4000	216
ORL	5	10304	50
Ear$_{64\times64}$	5	4096	30
Yale	5	77760	55



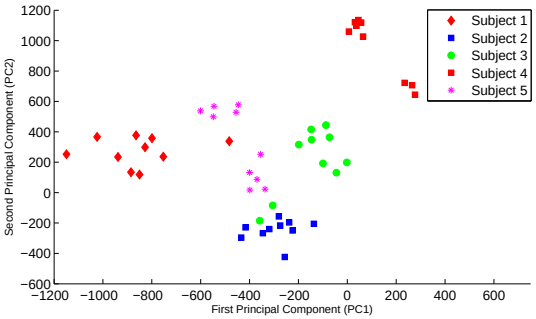
(a)



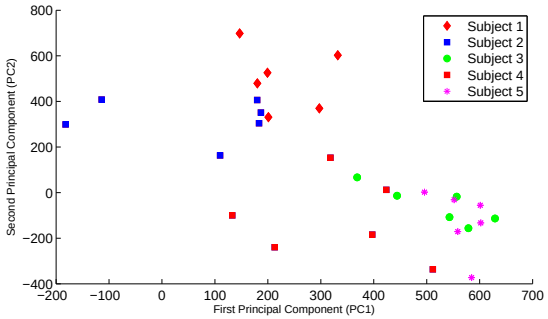
(b)



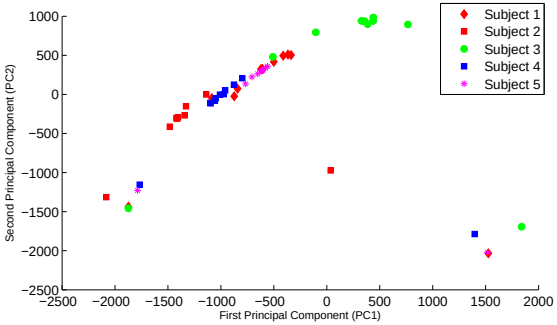
(c)



(d)

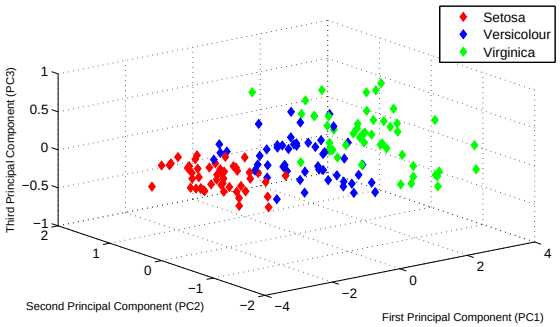


(e)

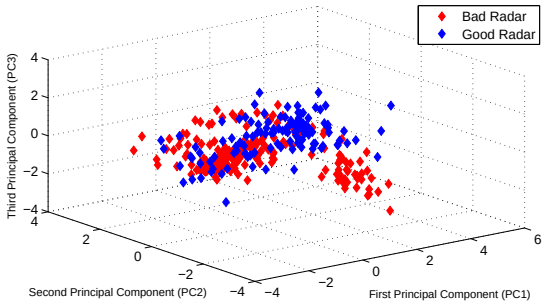


(f)

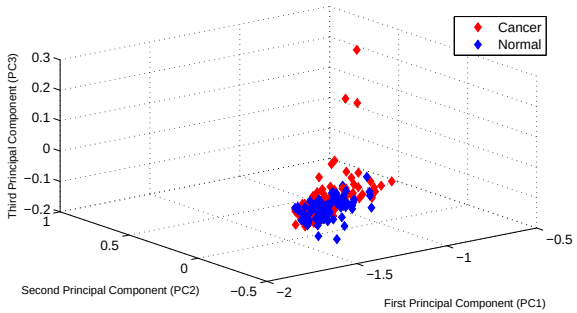
Figure: 2D Visualization of the datasets listed in Table 3, (a) Iris dataset, (b) Iono dataset, (c) Ovarian dataset, (d) ORL dataset, (e) Ear dataset, (f) Yale dataset.



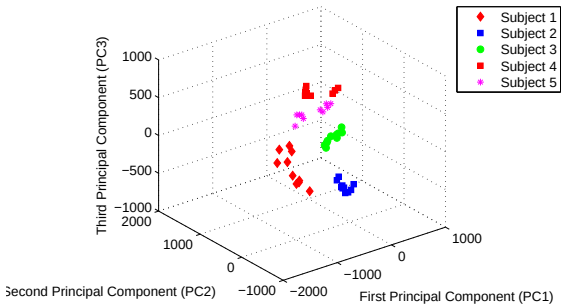
(a)



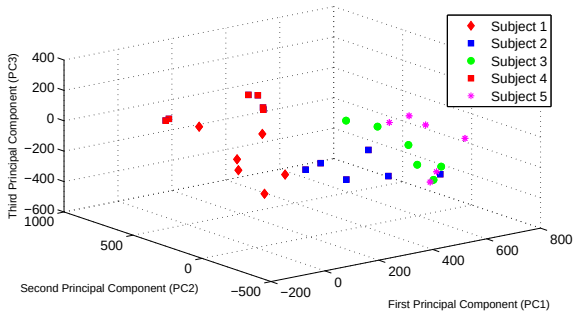
(b)



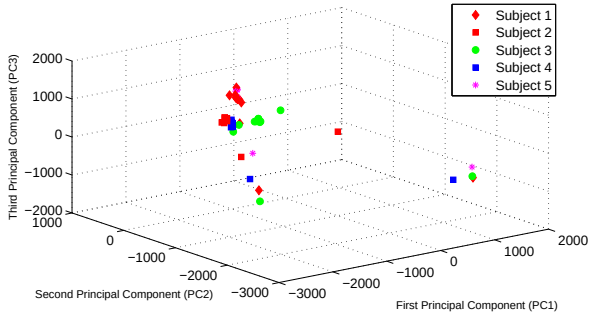
(c)



(d)



(e)



(f)

Figure: 3D Visualization of the datasets that were listed in Table 3, (a) Iris dataset, (b) Iono dataset, (c) Ovarian dataset, (d) ORL dataset, (e) Ear dataset, (f) Yale dataset.

Table: A comparison between 2D and 3D visualization in terms of MSE and robustness using the datasets that were listed in Table 3.

Dataset	2D		3D	
	Robustness (in %)	MSE	Robustness (in %)	MSE
Iris	97.76	0.12	99.48	0.05
Iono	43.62	0.25	51.09	0.23
Ovarian	98.75	0.04	99.11	0.03
ORL	34.05	24.03	41.64	22.16
Ear$_{64\times64}$	41.17	15.07	50.71	13.73
Yale	48.5	31.86	57.86	28.80

- Introduction to Principal component analysis (PCA)
  - Principal Components (PCs)
  - Covariance Matrix Method
  - Singular Value Decomposition (SVD) Method
  - PCA Space (Lower Dimensional Space)
  - Data Reconstruction
  - PCA Algorithms
- Numerical example
  - First Example: 2D-Class Example
  - Multi-Class Example
- Experimental Results and Discussion
  - Biometric Experiment
  - Image Compression Experiment
  - Data Visualization Experiment
- Conclusions

- For more details, read the original paper "Alaa Tharwat. "Principal component analysis-a tutorial" International Journal of Applied Pattern Recognition 3(3) (2016) 197-240"
- For more questions, send to engalaatharwat@hotmail.com