Dimensionality Reduction

Dr. Supaporn Erjongmanee

Department of Computer Engineering
Kasetsart University
fengspe@ku.ac.th

1

Outline

- Introduction
- Principal Component Analysis

2

Dimensionality Reduction

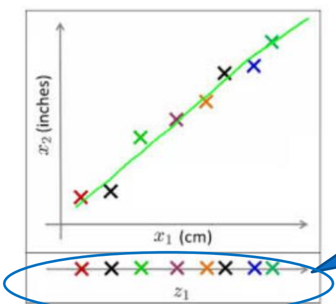
- Transformation process of data with many dimensions (many variables) to lower dimensions
- Two main approaches
  1. Feature selection
     - Selection subset of variables
  2. Feature extraction
     - Reduce higher dimensionality space to lower dimensionality subspace

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Feature Extraction Techniques

- Example:



We only need one dimension here.

- Three commonly-used techniques
  1. Principal Component Analysis (PCA)
  2. Linear Discriminant Analysis (LDA)
  3. Generalized Discriminant Analysis (GDA)

Image source: https://www.analyticsvidhya.com/blog/2015/07/dimension-reduction-methods/

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Feature Extraction Techniques (cont.)

- Advantages
  - Lower data storage
  - Lower computation time
- Disadvantages
  - Information loss
  - Meaning interpretation of lower dimension

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Review: Covariance Matrix

- Given random variables $X_1, X_2, \dots, X_p$
- Each entry in covariance matrix = $Cov(X_i, X_j)$
  - $Cov(X_i, X_j) = E[(X_i - E[X_i])(X_j - E[X_j])]$
- Note that $Cov(X_i, X_i) = \sigma^2_{X_i}$

$$\Sigma = \begin{bmatrix} \sigma^2_{X_1} & \cdots & Cov(X_1, X_p) \\ \vdots & \ddots & \vdots \\ Cov(X_p, X_1) & \cdots & \sigma^2_{X_p} \end{bmatrix}$$

Square (pxp) and Symmetric matrix

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Review: Eigenvectors and Eigenvalues

- Transformation of vector x in another space by changing its length (with the changing amount $= \lambda$), but not its direction

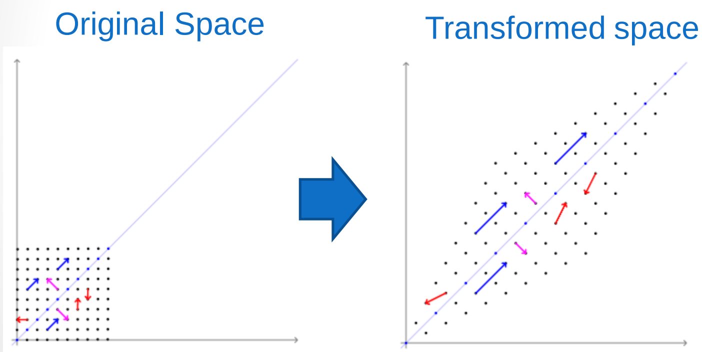


Image source: https://pathmind.com/wiki/eigenvector

Notice that red and blue vectors maintain the same direction, but not their lengths

Let A = Transformation matrix

$$\boxed{Ax = \lambda x}$$

Review: Eigenvectors and Eigenvalues

To find eigenvalues, we use $Ax = \lambda x$

$$(A - \lambda I)x = 0$$

A transforms vector x from one space to the other space

x is positive. For the above equation to have solution, $\det(A - \lambda I) = 0$

We solve for eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$

After deriving λ_i, use $Ae_i = \lambda_i e_i$ to solve for eigenvector e_i

Let $\Phi = [\boldsymbol{e_1} \quad \boldsymbol{e_2} \quad \cdots \quad \boldsymbol{e_p}]$, and $\Lambda = \begin{bmatrix} \lambda_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \lambda_p \end{bmatrix}$

We can rewrite $A = \Phi \Lambda \Phi^T$

Outline

- Introduction
- Principal Component Analysis

Principal Component Analysis

- Process of mapping data from higher dimension space to lower dimension subspace
- From p dimension $\Rightarrow$ q dimension

  p > q

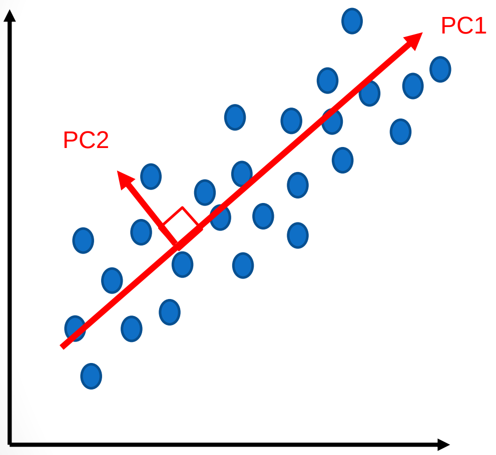
  Each dimension can be called component

Principal Component Analysis

- It is known that variable with large variance tends to explain output better than one with small variance
- Idea:
  1. Select the first component with largest variance.
  2. Then, select the next component that is orthogonal to the first one and also has large variance.
  3. Continue for all components
  4. Choose components that **most** explained original data (maybe with some information loss)

PCA Concept

- Example



- Choose principle components 1 (PC1) with largest variance
- Choose principle components 2 (PC2) orthogonal to PC1 and with large variance
- Do we need both PC1 and PC2?
- Is PC1 sufficient to explain data?

PCA Definition

- Let X be n samples, each sample has p variables

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \ddots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

- Let $\mu = [\mu_1, \mu_2, \ldots, \mu_p]$ = mean vector
- Let $\hat{X}$ = centered data matrix

$$\hat{X} = \begin{bmatrix} x_{11}-\mu_1 & x_{12}-\mu_2 & \cdots & x_{1p}-\mu_p \\ x_{21}-\mu_1 & x_{22}-\mu_2 & \cdots & x_{2p}-\mu_p \\ \cdots & \cdots & \ddots & \cdots \\ x_{n1}-\mu_1 & x_{n2}-\mu_2 & \cdots & x_{np}-\mu_p \end{bmatrix}$$

Note that
$E[\hat{X}] = E[X-\mu]$
$= E[X] - \mu = 0$

PCA Definition (cont.)

- Let Q = $\hat{X}^T \hat{X}$

$Q = \hat{X}^T \hat{X}$

$$= \begin{bmatrix} x_{11}-\mu_1 & x_{12}-\mu_2 & \cdots & x_{1p}-\mu_p \\ x_{21}-\mu_1 & x_{22}-\mu_2 & \cdots & x_{2p}-\mu_p \\ \cdots & \cdots & \ddots & \cdots \\ x_{n1}-\mu_1 & x_{n2}-\mu_2 & \cdots & x_{np}-\mu_p \end{bmatrix}^T \begin{bmatrix} x_{11}-\mu_1 & x_{12}-\mu_2 & \cdots & x_{1p}-\mu_p \\ x_{21}-\mu_1 & x_{22}-\mu_2 & \cdots & x_{2p}-\mu_p \\ \cdots & \cdots & \ddots & \cdots \\ x_{n1}-\mu_1 & x_{n2}-\mu_2 & \cdots & x_{np}-\mu_p \end{bmatrix}$$

$Q = \Sigma$ =Covariance Matrix

Square (pxp) and Symmetric

Projection Residuals

- Note that centered data have mean E[$\hat{X}$] = 0
- Let one centered data sample be $\vec{x_i}$
- Let $\vec{w}$ be *unit vector* of first component
- Then, the projected vector of $\vec{x_i}$ on first component will be $(\vec{w} \cdot \vec{x_i})\ \vec{w}$
- Projection residual = difference between the centered data and the projected vector = $\|\vec{x_i} - (\vec{w} \cdot \vec{x_i})\ \vec{w}\|^2$

PCA Setup

- From the p-dimensional space, find q-dimensional subspace
- Use data (p-dimensional vector) to project to different component

  Which component to choose?
- Goal: Each component will maximize variance
- Instead of finding component with maximum variance, we will show that it is equivalent to find component with smallest projection residual
  - To find component with maximum variance = To minimize projection residual

Minimize Projection Residuals (cont.)

$$\|\vec{x_i} - (\vec{w}\cdot\vec{x_i})\,\vec{w}\|^2 = ((\vec{x_i} - (\vec{w}\cdot\vec{x_i})\vec{w})\cdot(\vec{x_i} - (\vec{w}\cdot\vec{x_i})\vec{w})$$

$$= \vec{x_i}\cdot\vec{x_i} - \vec{x_i}\cdot(\vec{w}\cdot\vec{x_i})\vec{w} - (\vec{w}\cdot\vec{x_i})\vec{w}\cdot\vec{x_i} + (\vec{w}\cdot\vec{x_i})\vec{w}\cdot(\vec{w}\cdot\vec{x_i})\vec{w})$$

$$= \|\vec{x_i}\|^2 - 2(\vec{w}\cdot\vec{x_i})^2 + (\vec{w}\cdot\vec{x_i})^2\vec{w}\cdot\vec{w}$$

$$= \underbrace{\|\vec{x_i}\|^2 - (\vec{w}\cdot\vec{x_i})^2}_{\text{Residual of one data vector}} \qquad since\ \vec{w}\cdot\vec{w} = \|\vec{w}\|^2 = 1$$

Residual of one data vector

Average residuals of all n data vectors = MSE

17

Minimize Projection Residuals (cont.)

$$MSE(\vec{w}) = \frac{1}{n}\sum_{i=1}^{n}(\|\vec{x_i}\|^2 - (\vec{w}\cdot\vec{x_i})^2)$$

Average residuals of all n sample data vector

$$= \frac{1}{n}(\underbrace{\sum_{i=1}^{n}\|\vec{x_i}\|^2}_{\text{Do not depend on } \vec{w}} - \underbrace{\sum_{i=1}^{n}(\vec{w}\cdot\vec{x_i})^2}_{\text{To minimize MSE, make this term very large}})$$

Do not depend on $\vec{w}$ To minimize MSE, make this term very large

Hence, to minimize MSE, we want to maximize $\frac{1}{n}\sum_{i=1}^{n}(\vec{w}\cdot\vec{x_i})^2$

18

Minimize Projection Residuals (cont.)

$$Var[Y] = E[Y^2] - E[Y]^2 = \frac{1}{n}(\sum_{i=1}^{n} y_i^2) - (\frac{1}{n}\sum_{i=1}^{n} y_i)^2$$

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n}(\vec{w}\cdot\vec{x_i})^2} = \underbrace{(\frac{1}{n}\sum_{i=1}^{n}\vec{w}\cdot\vec{x_i})^2} + \underbrace{Var[\vec{w}\cdot\vec{x_i}]}$$

To minimize MSE, we maximize this ➡ Equal to zero (see below) ➡ Hence, to minimize MSE, we maximize this

$$\frac{1}{n}\sum_{i=1}^{n}(\vec{w}\cdot\vec{x_i}) = (\frac{1}{n}\vec{w}(\underbrace{\sum_{i=1}^{n}\hat{x}_i}) \cdot)$$

Note that $\hat{X}$ = centered data, $E[\hat{X}] = 0$

$$\|\vec{x_i} - (\vec{w}\cdot\vec{x_i})\,\vec{w}\|^2$$

To minimize projection residual
= To find projection with
maximum variance

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Minimize Projection Residuals (cont.)

$$Var[Y] = E[Y^2] - E[Y]^2 = \frac{1}{n}(\sum_{i=1}^{n} y_i^2) - (\frac{1}{n}\sum_{i=1}^{n} y_i)^2$$

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n}(\vec{w}\cdot\vec{x_i})^2} = \underbrace{(\frac{1}{n}\sum_{i=1}^{n}\vec{w}\cdot\vec{x_i})^2} + \underbrace{Var[\vec{w}\cdot\vec{x_i}]}$$

To minimize MSE, we maximize this ➡ Equal to zero (see below) ➡ Hence, to minimize MSE, we maximize this

This is variance of projection to first principal component

We need to project all $\vec{x_i}'s$ to other principal components.

Hence, to minimize MSE = to maximize sum of variances to all components

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Maximize Projection Variance

- Note that $\hat{\boldsymbol{x}}$ be n centered data vectors with p dimension
  - $\hat{\boldsymbol{x}} = n \times p$ matrix
- Given $\vec{w}$ be *unit vector* of each projection, then projection of all $\hat{\boldsymbol{x}}$ onto all projections = $\hat{\boldsymbol{x}}\boldsymbol{w}$
  - $\hat{\boldsymbol{x}}\boldsymbol{w} = n \times 1$

$$\underbrace{\sigma^2_{\vec{w}} = \frac{1}{n}\sum_{i=1}^{n}(\vec{w}\cdot\vec{x_i})^2}_{\text{Variance of one projection Using n data}} = \frac{1}{n}(\hat{\boldsymbol{x}}\boldsymbol{w})^T(\hat{\boldsymbol{x}}\boldsymbol{w}) = \frac{1}{n}\boldsymbol{w}^T\hat{\boldsymbol{x}}^T\hat{\boldsymbol{x}}\boldsymbol{w}$$

$$= \boldsymbol{w}^T\frac{\hat{\boldsymbol{x}}^T\hat{\boldsymbol{x}}}{n}\boldsymbol{w} = \boldsymbol{w}^T\boldsymbol{\Sigma}\boldsymbol{w}$$

Find w that maximizes $\sigma^2_{\vec{w}}$

$\boldsymbol{\Sigma} = \hat{\boldsymbol{x}}^T\hat{\boldsymbol{x}}$ ($p \times p$ matrix)

Supaporn Erjongmanee
fengspe@ku.ac.th

21

Maximize Projection Variance (cont.)

- Maximize $\sigma^2_{\vec{w}}$ with constraint $\boldsymbol{w}^T\ \boldsymbol{w}$ = 1
- Solve by using Lagrange Multiplier

$$\mathcal{L}(\boldsymbol{w},\lambda) = \sigma^2_{\boldsymbol{w}} - \lambda(\boldsymbol{w}^T\boldsymbol{w} - \boldsymbol{1}) = \boldsymbol{w}^T\boldsymbol{v}\,\boldsymbol{w} - \lambda(\boldsymbol{w}^T\boldsymbol{w} - \boldsymbol{1})$$

$$\frac{\partial\mathcal{L}}{\partial\lambda} = \boldsymbol{w}^T\boldsymbol{w} - \boldsymbol{1} \xrightarrow{\text{Set to zero}} \boldsymbol{w}^T\boldsymbol{w} = \boldsymbol{1}$$

$$\frac{\partial\mathcal{L}}{\partial w} = 2\boldsymbol{\Sigma}\boldsymbol{w} - \boldsymbol{2}\lambda\boldsymbol{w} \xrightarrow{\text{Set to zero}} \boldsymbol{\Sigma}\boldsymbol{w} = \lambda\boldsymbol{w}$$

w that maximizes $\sigma^2_{\vec{w}}$ = eigenvector of $\boldsymbol{\Sigma}$ with eigenvalue = λ

Eigenvector of covariance matrix

Supaporn Erjongmanee
fengspe@ku.ac.th

22

Principal Components

$\Sigma w = \lambda w$

- Since Σ is $p \times p$ matrix, there will be p eigenvectors
- Σ is covariance matrix ->Σ is symmetric
  - Eigenvectors are orthogonal to each other
- Σ is covariance matrix -> values in Σ are positive
  - Eigenvalues are positive
- p eigenvectors of Σ= p principal components of $\hat{x}$

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Principal Components (cont.)

- p Eigenvectors of Σ= p principal components of $\hat{x}$
  - Eigenvector with largest eigenvalue = e_1= first principal component
  - Eigenvector with second largest eigenvalue = e_2 = second principal component
  - …
  - Eigenvector with the pth largest eigenvalue = e_p = pth principal component
- Note that each principal component is orthogonal to each other
- Each eigenvalue = variance described by each component
- Sum of eigenvalues = total variances described by all components

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Principal Components (cont.)

- There are p eigenvectors $\boldsymbol{e_1}, \boldsymbol{e_1}, \ldots, \boldsymbol{e_p}$ that correspond to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$

$$\boldsymbol{\Sigma e_1} = \lambda \boldsymbol{e_1}, \boldsymbol{\Sigma e_2} = \lambda \boldsymbol{e_2}, \ldots, \boldsymbol{\Sigma e_p} = \lambda \boldsymbol{e_p}$$

$$\underbrace{\boldsymbol{\Sigma}[\boldsymbol{e_1} \quad \boldsymbol{e_2} \quad \cdots \quad \boldsymbol{e_p}]}_{\Phi} = [\boldsymbol{e_1} \quad \boldsymbol{e_2} \quad \cdots \quad \boldsymbol{e_p}] \underbrace{\begin{bmatrix} \lambda_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \lambda_p \end{bmatrix}}_{\Lambda}$$

$$\boxed{\boldsymbol{\Sigma}\Phi = \Phi\Lambda}$$

25

Principal Components (cont.)

- If we normalize eigenvector to unit vector

$$\Phi\Phi^T = \Phi^T\Phi = I$$

$$\boxed{\boldsymbol{\Sigma}\Phi = \Phi\Lambda} \Rightarrow \Phi^T\boldsymbol{\Sigma}\Phi = \Lambda$$

$$\boldsymbol{\Sigma} = \Phi\Lambda\Phi^T$$

Eigenvector matrix Eigenvalue matrix

26

PCA Procedure

1. Compute $\hat{X}$ — Dimension of X = [n x p]
2. Compute $\mathbf{\Sigma}$ = covariance matrix of $\hat{X}$
3. Compute single value decomposition (SVD) of $\mathbf{\Sigma} = \Phi\Lambda\Phi^T$
   - Λ = diag($\lambda_1, \lambda_2, \ldots, \lambda_p$) where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$
   - $\Phi = [e_1, e_2, \ldots, e_p]$
4. Choose q < p and obtain $\Phi_q = [e_1, e_2, \ldots, e_q]$
5. Obtain $\boldsymbol{y} = \Phi_q^T \boldsymbol{x}$ — Dimension of Y = [n x q]

Question: How to choose q?

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

27

How to choose q?

- $\sum_{i=1}^{p} \lambda_i$ = total variance described by all components
- $\sum_{i=1}^{q} \lambda_i$ = variance explained by PCA
- Define fraction of original variance and input vectors
  - $R^2 = \frac{\sum_{i=1}^{q} \lambda_i}{\sum_{i=1}^{p} \lambda_i} \geq 1 - \alpha$
  - where α = accepted error and $0 < \alpha < 1$

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

28

Example

- Beside variances, eigenvalues also specify percent of transformation along each component
- If there are 3 eigenvectors with eigenvalues
  - $\{\lambda_1 = 3,\ \lambda_2 = 2,\ \lambda_3 = 1\}$
  - Percent of transformation occurs in 1st component = 50%
  - Percent of transformation occurs in 2nd component = 33.33%
  - Percent of transformation occurs in 3rd component = 16.67%
- If we choose only 2 principal components (q=2),
  - Percent of data explained by PCA = 83.33%
  - Some information will be lost. Question: Is it acceptable?

Projection Residuals vs. q

- Let Φ_q = q eigenvectors
  - Projection residual = $x - \Phi_q x$
- If data are really q-dimensional, there are q eigenvectors with q positive eigenvalues.
  - The remaining p-q eigenvalues ≈ 0
  - Projection residual = 0
- If data approximately have q-dimensions, projection residual is small
- If data is larger than q-dimensions, projection residual is large

Example: Data Projection

- Obtain $y_i = \Phi_q^T x_i$
- Let p = 3, q = 3

No dimensionality reduction

$$x_i = \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ x_{i,3} \end{bmatrix}$$

$$\Phi = [\boldsymbol{e_1} \quad \boldsymbol{e_2} \quad \boldsymbol{e_3}]$$

$$\Phi^T x_i = \begin{bmatrix} \boldsymbol{e_1} \\ \boldsymbol{e_2} \\ \boldsymbol{e_3} \end{bmatrix} \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ x_{i,3} \end{bmatrix} = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ x_{i,3} \end{bmatrix} = \begin{bmatrix} y_{i,1} \\ y_{i,2} \\ y_{i,3} \end{bmatrix} = y_i$$

Example 2: Data Projection

- Obtain $y_i = \Phi_q^T x_i$
- Let p = 3, q = 2

Dimensionality reduction

$$x_i = \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ x_{i,3} \end{bmatrix}$$

$$\Phi = [\boldsymbol{e_1} \quad \boldsymbol{e_2} \quad \boldsymbol{e_3}]$$

$$\Phi_q = [\boldsymbol{e_1} \quad \boldsymbol{e_2}]$$

$$\Phi_q^T x_i = \begin{bmatrix} \boldsymbol{e_1} \\ \boldsymbol{e_2} \end{bmatrix} \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ x_{i,3} \end{bmatrix} = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \end{bmatrix} \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ x_{i,3} \end{bmatrix} = \begin{bmatrix} y_{i,1} \\ y_{i,2} \end{bmatrix} = y_i$$

Example: Data Reconstruction

- *Reconstruct* $x_i = \Phi_q \, y_i$
- Let p = 3, q = 3

$$y_i = \begin{bmatrix} y_{i,1} \\ y_{i,2} \\ y_{i,3} \end{bmatrix}$$

$$x_i = \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ x_{i,3} \end{bmatrix}$$

$$\Phi = [\boldsymbol{e_1} \quad \boldsymbol{e_2} \quad \boldsymbol{e_3}]$$

$$\Phi_q \, y_i = [\boldsymbol{e_1} \quad \boldsymbol{e_2} \quad \boldsymbol{e_3}] \begin{bmatrix} y_{i,1} \\ y_{i,2} \\ x_{i,3} \end{bmatrix} = \begin{bmatrix} e_{11} & e_{21} & e_{31} \\ e_{12} & e_{22} & e_{32} \\ e_{13} & e_{23} & e_{33} \end{bmatrix} \begin{bmatrix} y_{i,1} \\ y_{i,2} \\ y_{i,3} \end{bmatrix} = \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ x_{i,3} \end{bmatrix} = x_i$$

No information loss

Example 2: Data Reconstruction

- *Reconstruct* $x_i = \Phi_q \, y_i$
- Let p = 3, q = 2

$$y_i = \begin{bmatrix} y_{i,1} \\ y_{i,2} \end{bmatrix}$$

$$x_i = \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ x_{i,3} \end{bmatrix}$$

$$\Phi = [\boldsymbol{e_1} \quad \boldsymbol{e_2} \quad \boldsymbol{e_3}]$$

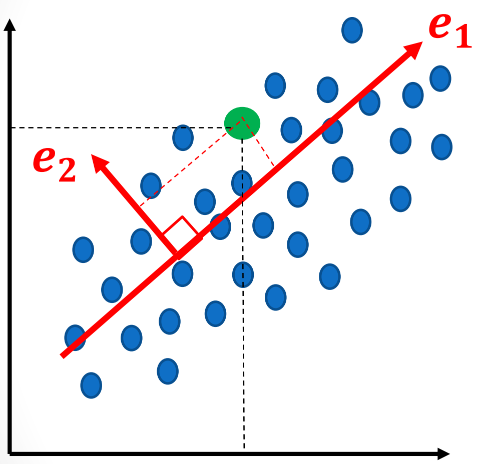
$$\Phi_q = [\boldsymbol{e_1} \quad \boldsymbol{e_2}]$$

$$\Phi_q \, y_i = [\boldsymbol{e_1} \quad \boldsymbol{e_2}] \begin{bmatrix} y_{i,1} \\ y_{i,2} \end{bmatrix} = \begin{bmatrix} e_{11} & e_{21} \\ e_{12} & e_{22} \\ e_{13} & e_{23} \end{bmatrix} \begin{bmatrix} y_{i,1} \\ y_{i,2} \end{bmatrix} = \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ x_{i,3} \end{bmatrix} = x_i$$

Some information loss

PCA Transformation

- Visualization



$$\Sigma = \Phi\Lambda\Phi^T$$

Calculate eigenvectors of covariance matrix Σ

$\Phi = [e_1, e_2]$

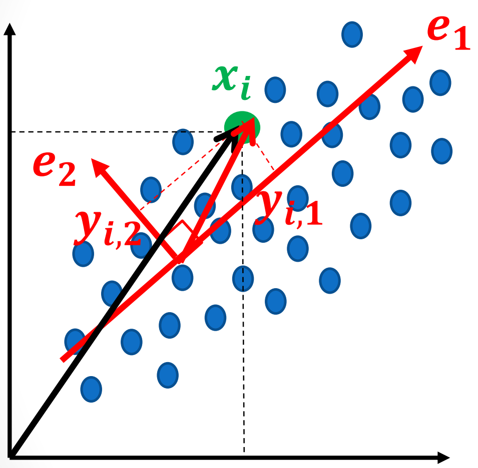
We derive PC1 = e_1 ,
PC2 = e_2

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Transformation

- Visualization



We derive
PC1 = e_1
PC2 = e_2

We can project each data point to principal component space

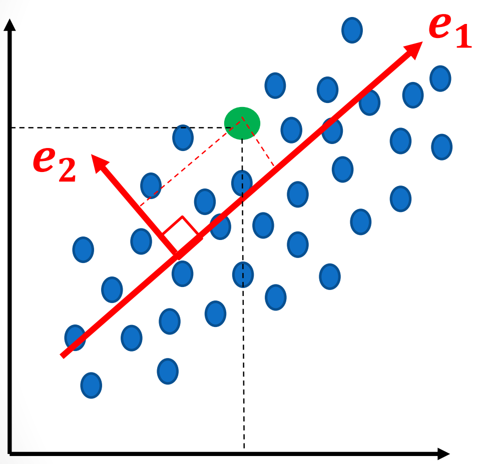
$$y_i = [e_1, e_2]^T \overrightarrow{x_i} = \Phi^T \overrightarrow{x_i}$$

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Transformation

- Visualization



With PC1 = $\boldsymbol{e_1}$, PC2 = $\boldsymbol{e_2}$

We can reconstruct each data point from principal components

$$\tilde{x}_i = [\boldsymbol{e_1}, \boldsymbol{e_2}]y_i$$

The reconstructed $\tilde{x}_i$ may not have the same information as $\vec{x_i}$ if q < p

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

Conclusion

- Dimensionality reduction is to reduce higher dimensional space to lower dimensional subspace
  - Some original information may be lost
- Principal component analysis is to find new components of data where each component span over maximum variances and orthogonal to other component
  - Some component with small eigenvalues (variances) can be ignored
  - This allows us to reduce data dimensionality

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University

References

1. C. Shalizi, Principal Component Analysis, available at https://www.stat.cmu.edu/~cshalizi/uADA/12/lectures/ch18.pdf
2. L. Wiskott, Lecture Notes on Principal Component Analysis, available at http://cs233.stanford.edu/ReferencedPapers/LectureNotes-PCA.pdf
3. P. Rigollet, Principal Component Analysis, Statistics for Applications, available at: https://ocw.mit.edu/courses/mathematics/18-650-statistics-for-applications-fall-2016/lecture-slides/MIT18_650F16_PCA.pdf

Supaporn Erjongmanee
fengspe@ku.ac.th

Department of Computer Engineering
Kasetsart University