# Principal Component Analysis

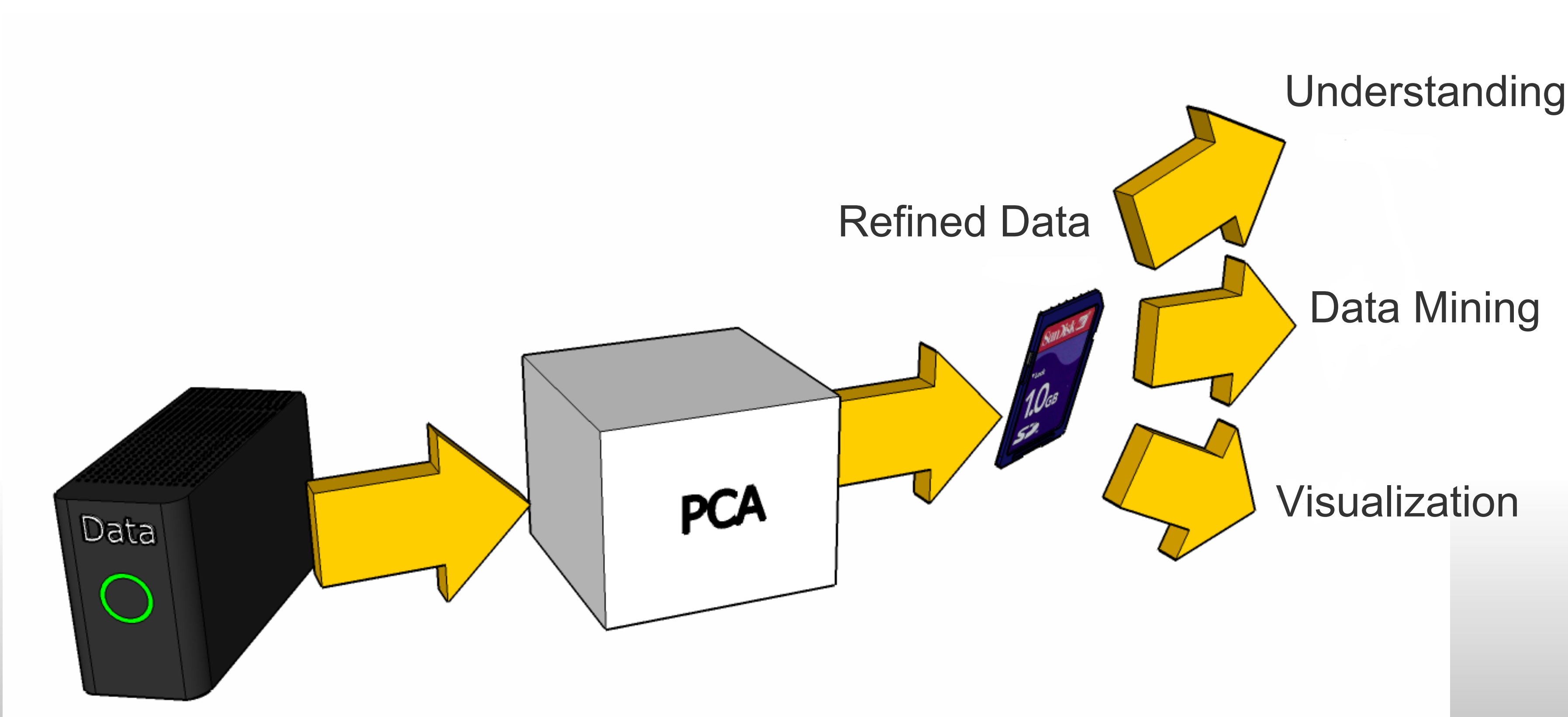
Morgan Bengtsson benmo417@student.liu.se
Cliver Zardán cliza724@student.liu.se

# Introduction: In this lecture

- What is PCA?
- Example: The Spring
- Mathematical basics
  - Linear Algebra
  - Statistics
- Example Continued
  - Steps
- Data mining and PCA
- Conclusions
- References

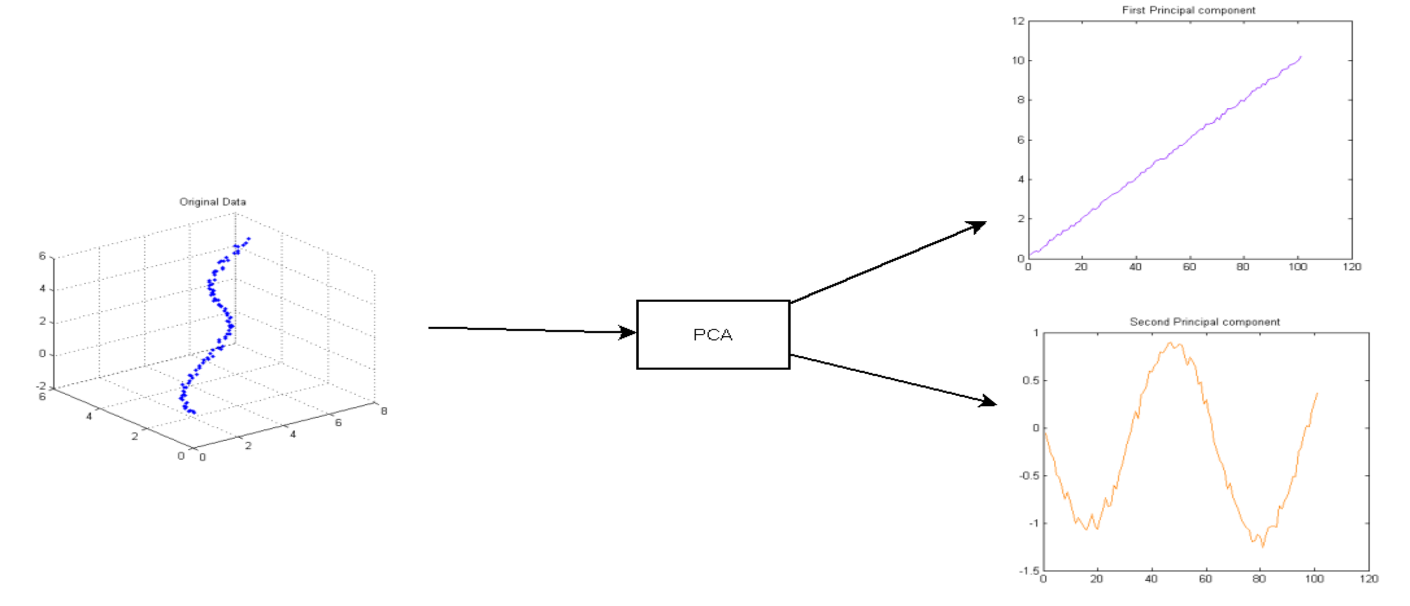
# What is PCA?

- Magic Box



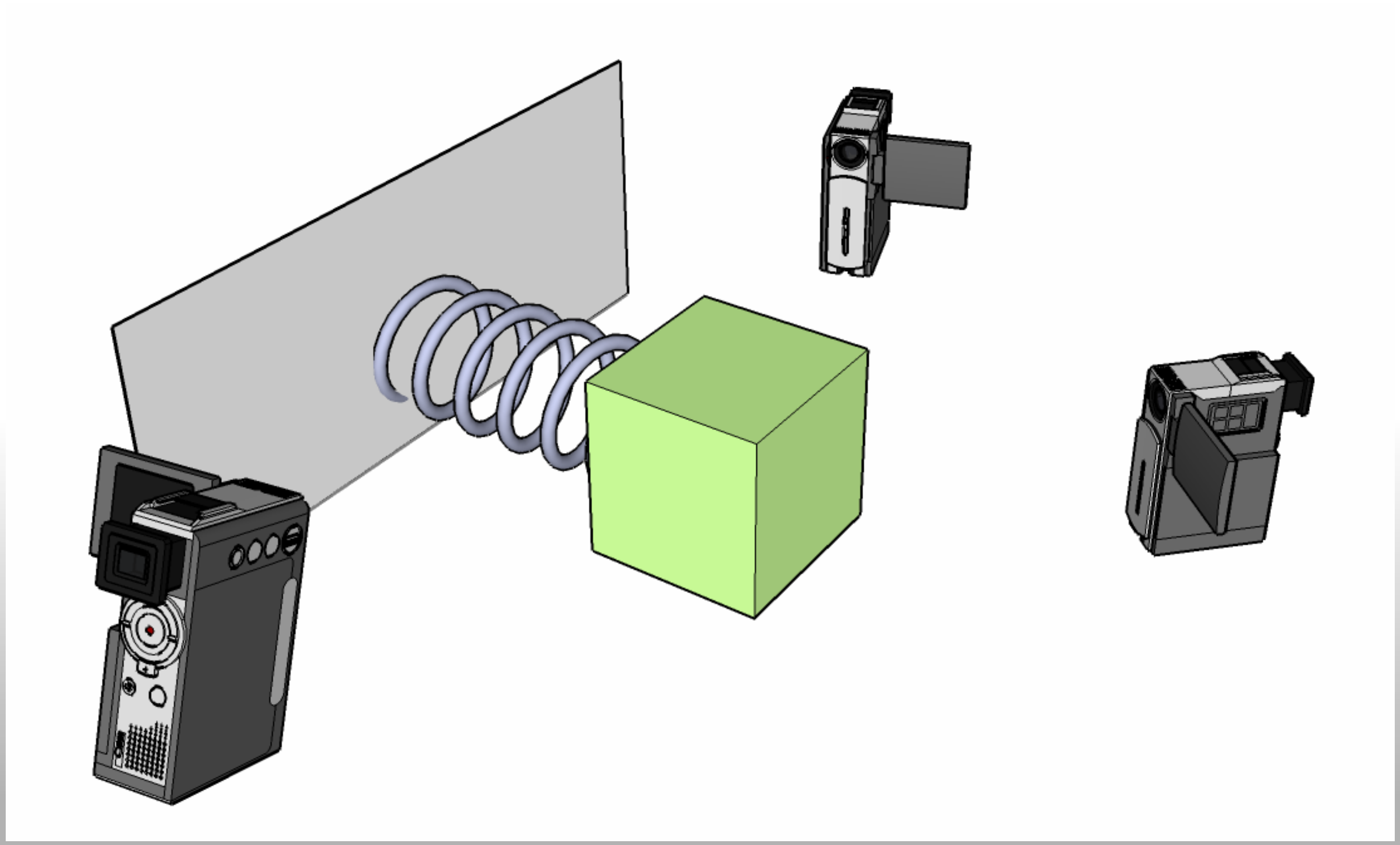
# What is PCA?

- Extract of relevant information
- Find patterns
- Relatively simple
- Reduce of dimensions
- Change of basis
- Somewhat remove noise



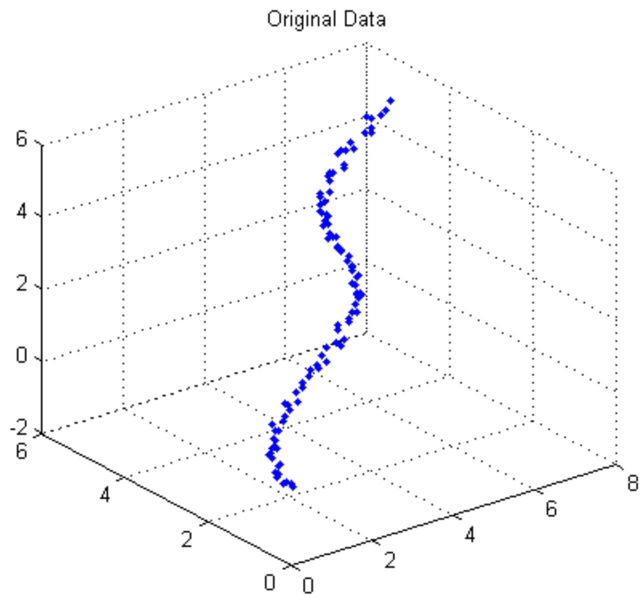
# Example: The Spring

- Artificially created
- Could be relatively real though
- Ideal spring



# Example: The Spring Data

- Complicated multidimensional dataset




| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | 0.1231 | 0.2655 | 0.3258 | 0.3411 | 0.3295 | 0.3783 | 0.5683 |
| 2 | 0.1195 | 0.1767 | 0.2925 | 0.4223 | 0.5193 | 0.6965 | 0.887 |
| 3 | 0.137 | 0.0142 | 0.0345 | -0.0205 | -0.0595 | 0.0488 | -0.0976 |

# Matematical Basics

Linear algebra:

- Eigen vectors
- Eigen values
- Matrix Algebra

Statistics:

- Standard Deviation
- Variance
- Covariance
- Covariance Matrix

# Linear Algebra: Eigenvalues & Eigenvectors

Eigenvectors can only be found for square matrices.

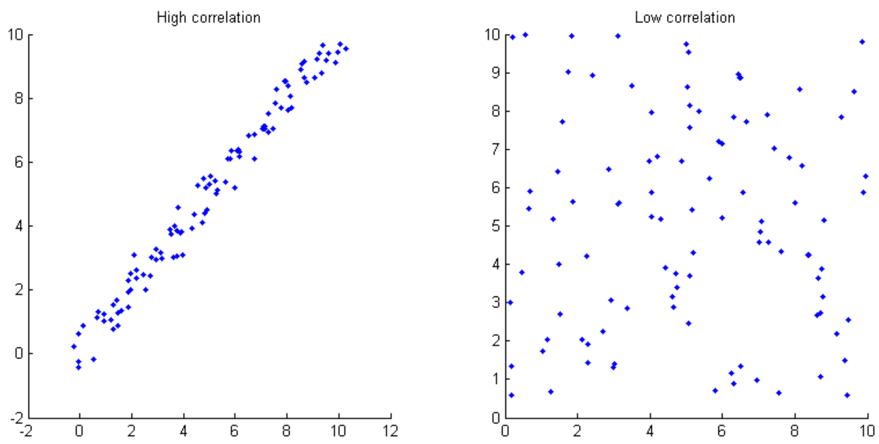
Eigenvectors and eigenvalues comes in pairs.

Orthogonal

Ex. A 3x3 matrix has 3 eigenvectors. The highest eigenvalue represents the best eigenvector.

# Statistics: Covariance and Covariance Matrix

Covariance = Dependence betwen two sets



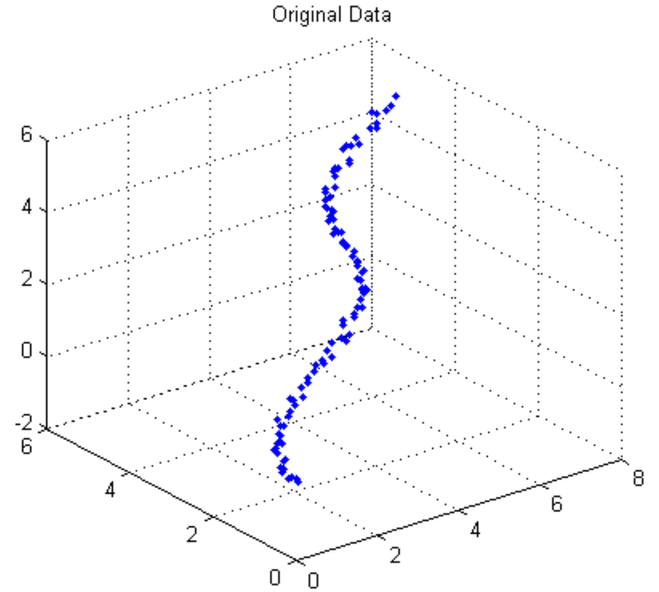

Covariance Matrix = If working with many dimensions

$$C = \begin{pmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{pmatrix}$$

Covariance Matrix of a 3-dimensional data set.

# Example Continued

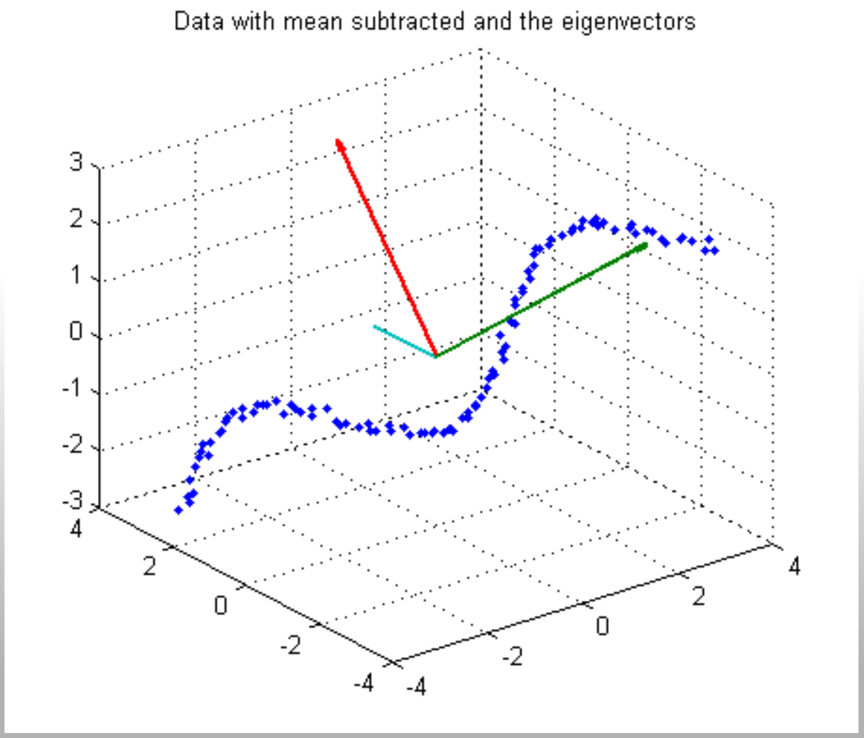
- Complicated multidimensional dataset




| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | 0.1231 | 0.2655 | 0.3258 | 0.3411 | 0.3295 | 0.3783 | 0.5683 |
| 2 | 0.1195 | 0.1767 | 0.2925 | 0.4223 | 0.5193 | 0.6965 | 0.887 |
| 3 | 0.137 | 0.0142 | 0.0345 | -0.0205 | -0.0595 | 0.0488 | -0.0976 |

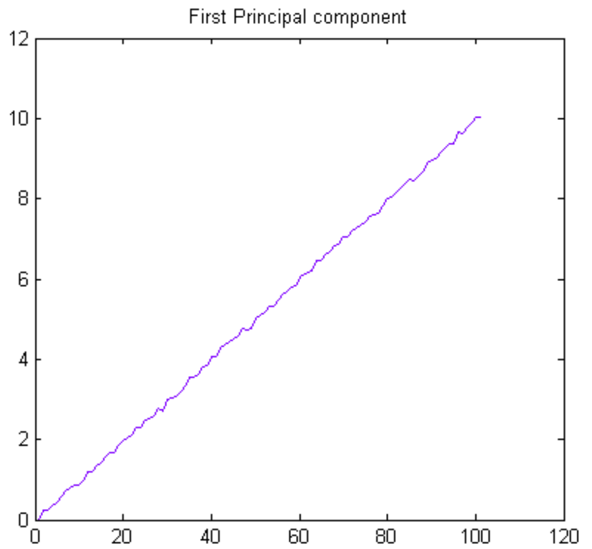
# Example Continued: PCA Steps 1 - 7

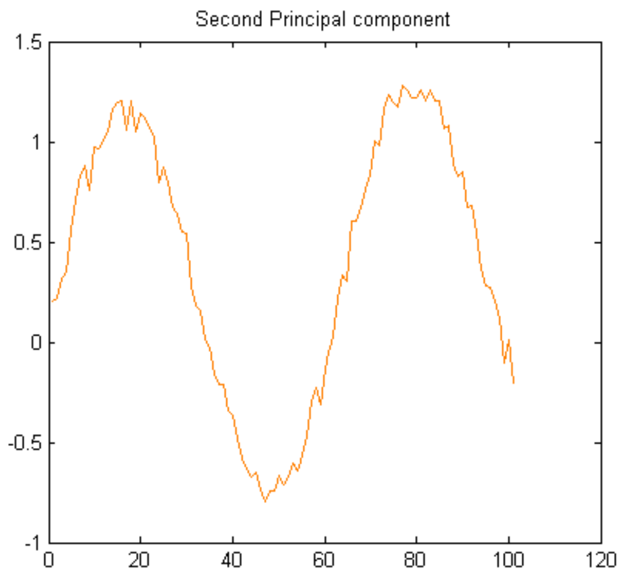
1. Subtract the mean along each dimension
2. Calculate the covariance matrix
    - cov = 1/(N-1) *dataDist*dataDist';
1. Calculate the *Principal Components* of cov. (eigs in Matlab)
    - Sorted with respect to eigenvalues

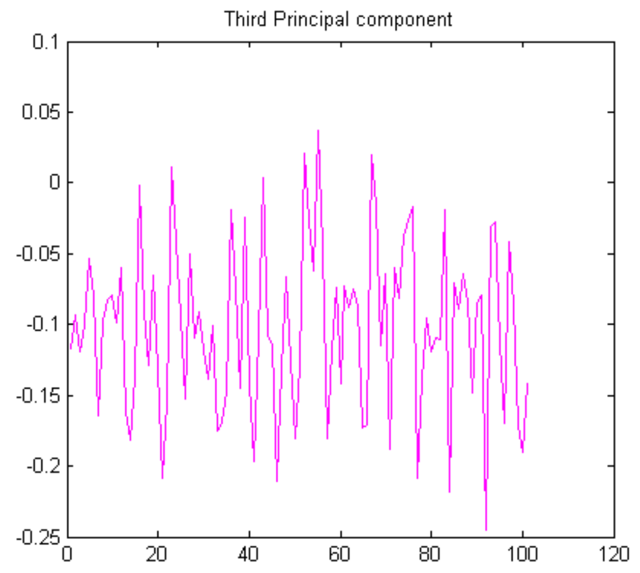


# Example Continued: PCA Steps 1 - 7

4. Select the principal components that are relevant
   - Three principal components
   - Data transformed with respect to each PC

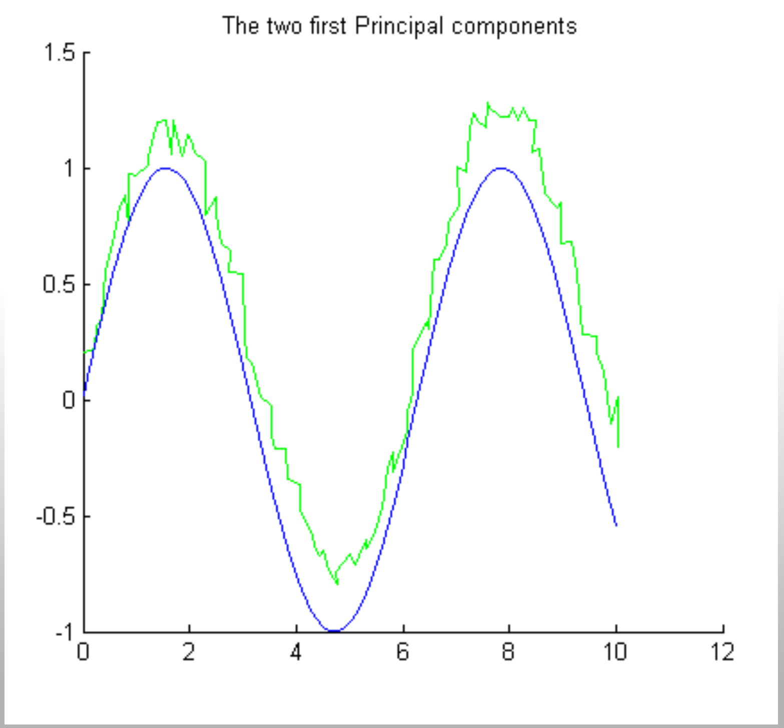


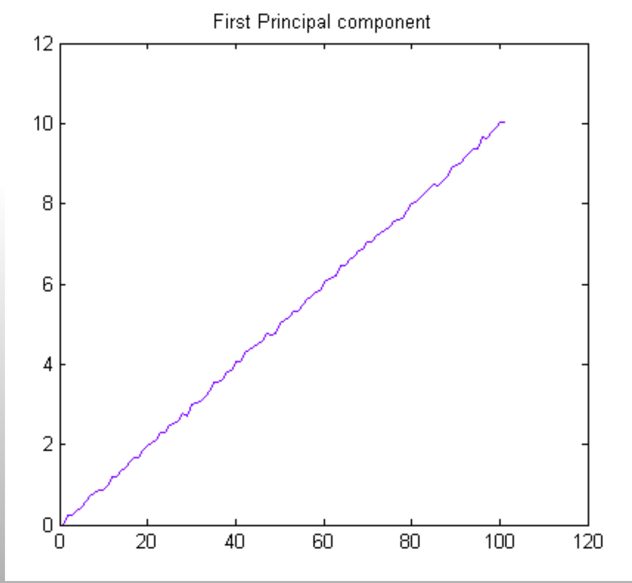
# Example Continued: PCA Steps

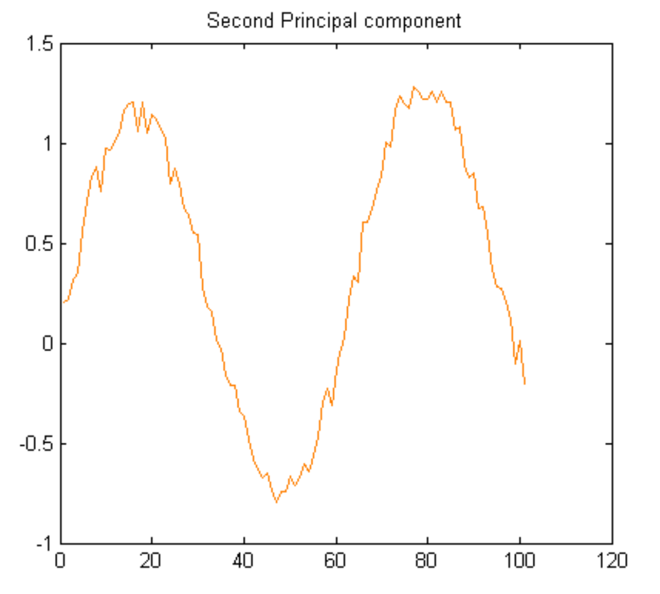
5. Transform with respect to those (Feature Vector)
   - dataPC12 = PC(:,1:2)'*dataDist;
6. Choose
   - Transform to original coordinates
   - Keep the new coordinate system
7. Re add the mean values

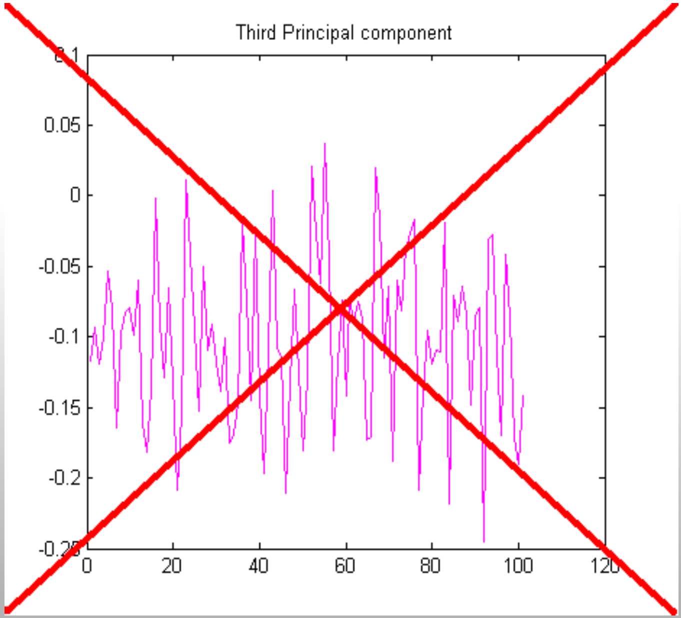


# Example: Summary

- Complicated data set
- Found principal components
- Reduced dimension
- Transformed to new basis



# Data mining and PCA

- Better representation of data
- More representative basis
- Accuracy of classification model
- Faster and better data processing

# Data mining and PCA

- Applications
  - Military
  - Medicine
  - Experiments
  - Neuroscience
  - Computer Graphics
  - Infovis
  - ...

# Conclucions

Strengths

- Easy
- Widely used
- Efficent
- No tweaking

Weaknesses

- No tweaking

# References


Principal Components

http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

A Tutorial on Principal Component Analysis

http://www.snl.salk.edu/%7Eshlens/pub/notes/pca.pdf

Principal Components Analysis

http://www.resample.com/xlminer/help/PCA/pca_intro.htm

Questions?