MACHINE LEARNING FOR SIGNAL PROCESSING

8-1-2025

Sriram Ganapathy
LEAP lab, Electrical Engineering, Indian Institute of Science,
sriramg@iisc.ac.in

Viveka Salinamakki, Varada R.
LEAP lab, Electrical Engineering, Indian Institute of Science.
viveka.sg@gmail.com varadar2000@gmail.com

http://leap.ee.iisc.ac.in/sriram/teaching/MLSP25/

VECTOR CALCULUS

- Let $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ denote a real n dimensional vector $\in \mathscr{R}^{n\times 1}$ and let $f(\boldsymbol{x})$ denote a scalar function which maps $\mathscr{R}^n \to \mathscr{R}$
- Then, we define $\frac{\partial f}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$ is derivate of the function $f(\boldsymbol{x})$ w.r.t. $\boldsymbol{x}$
- Question :
  - ✴ What is the derivative of $f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x}$
  - ✴ What is the derivative of $f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{x}$

VECTOR DERIVATIVES

- Let $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ denote a real n dimensional vector $\in \mathscr{R}^{n\times 1}$ and let $\boldsymbol{f}(\boldsymbol{x})$ denote a vector function which maps $\mathscr{R}^n \to \mathscr{R}^m$

… Let $\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ f_2(\boldsymbol{x}) \\ \vdots \\ f_m(\boldsymbol{x}) \end{bmatrix}$

- Then, we define $\dfrac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & & & \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$ is derivate of the function $\boldsymbol{f}(\boldsymbol{x})$ w.r.t. $\boldsymbol{x}$ … The derivative $\in \mathscr{R}^{m\times n}$

MATRIX DERIVATIVES

- Let $A = \begin{bmatrix} a_{11}a_{12}..a_{1n} \\ a_{21}a_{22}..a_{2n} \\ \vdots \\ a_{m1}a_{m2}..a_{mn} \end{bmatrix}$ denote a real m x n matrix and let $f(A)$ denote a scalar function which maps $\mathscr{R}^{m\times n} \to \mathscr{R}$
- Then, we define $\dfrac{\partial f}{\partial A} = \begin{bmatrix} \dfrac{\partial f}{\partial a_{11}} \dfrac{\partial f}{\partial a_{12}} .. \dfrac{\partial f}{\partial a_{1n}} \\ \dfrac{\partial f}{\partial a_{21}} \dfrac{\partial f}{\partial a_{22}} .. \dfrac{\partial f}{\partial a_{2n}} \\ \vdots \\ \dfrac{\partial f}{\partial a_{m1}} \dfrac{\partial f}{\partial a_{m2}} .. \dfrac{\partial f}{\partial a_{mn}} \end{bmatrix}$ is derivate of the function $f(A)$ w.r.t. A …
- This derivative $\in \mathscr{R}^{m\times n}$

MATRIX DERIVATIVE PROBLEMS

➤ Question :

- What is the derivative of $f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x}$ w.r.t $\boldsymbol{x}$
- What is the derivative of $f(\boldsymbol{A}) = \boldsymbol{x}^T \boldsymbol{A}\boldsymbol{x}$ w.r.t $\boldsymbol{A}$
- What is the derivative of $f(\boldsymbol{A}) = Tr(\boldsymbol{A})$ w.r.t $\boldsymbol{A}$
- What is the derivative of $f(\boldsymbol{A}) = Tr(\boldsymbol{A}\boldsymbol{B})$ w.r.t $\boldsymbol{A}$
- What is the derivative of $f(\boldsymbol{A}) = Tr(\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}^T)$ w.r.t $\boldsymbol{A}$
- What is the derivative of $f(\boldsymbol{A}) = log\,|\boldsymbol{A}|$ w.r.t $\boldsymbol{A}$

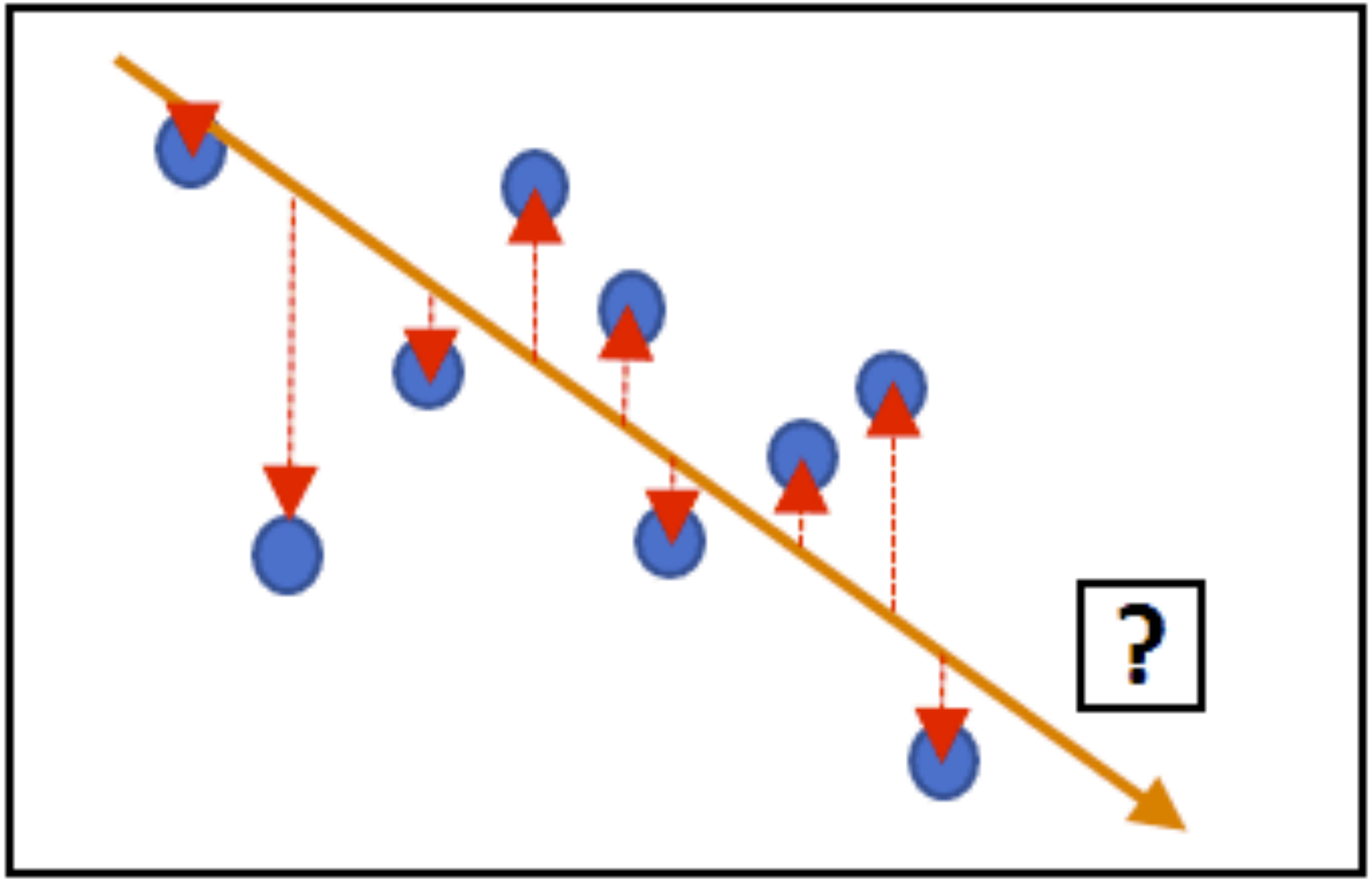
INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

LEAP

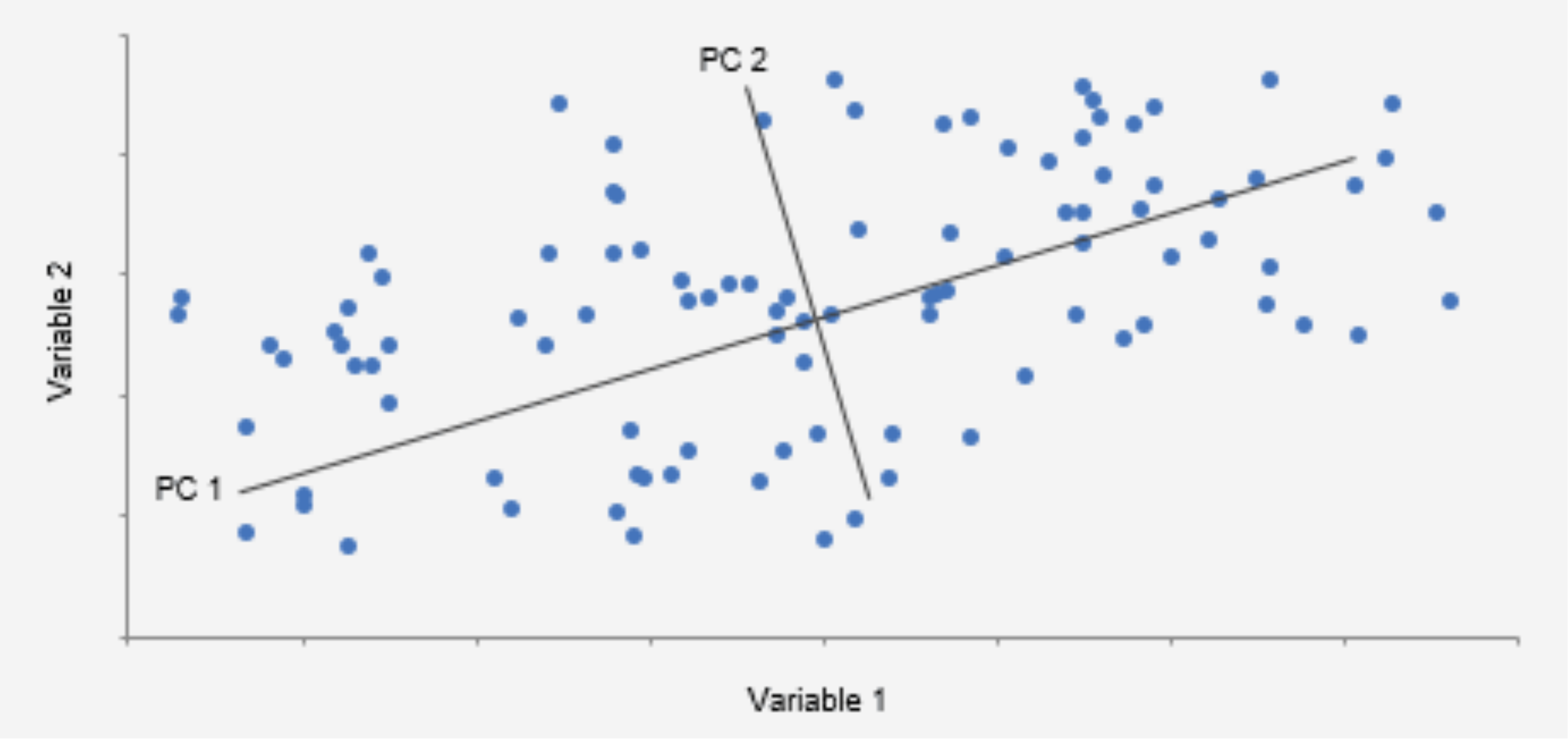
PRINCIPAL COMPONENT ANALYSIS

- Reducing the data $\mathbf{x}_n$ of dimension D to lower dimension
- Projecting the data into subspace which preserves maximum data variance
  - ✓ Maximize variance in projected space $M < D$
- Equivalent formulated as minimizing the error between the original and projected data points.

DIRECTION OF MAXIMUM VARIANCE



PCA EXAMPLE



PRINCIPAL COMPONENT ANALYSIS

- First M eigenvectors of data covariance matrix

$$S = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T$$

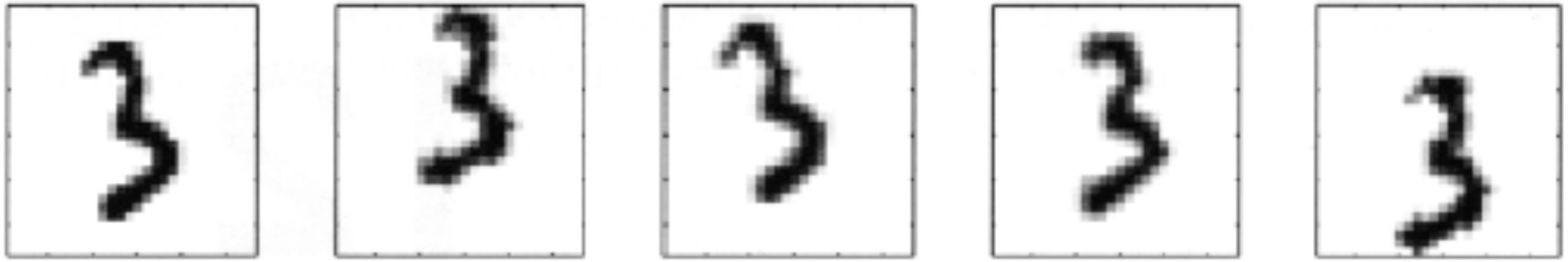
- Residual error from PCA

$$J = \sum_{i=M+1}^{D} \lambda_i$$

PRML - C. Bishop (Sec. 12.1)

PCA

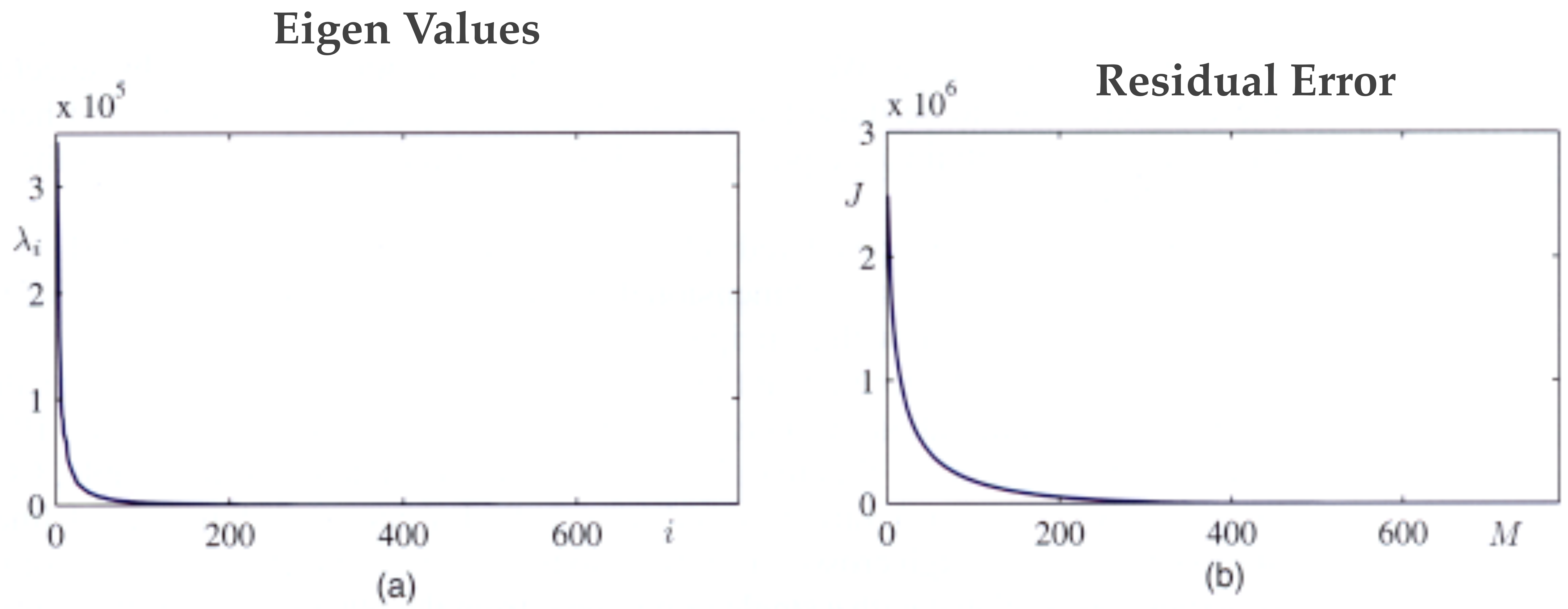
- First   eigenvectors of data covariance matrix



- Residual error from PCA

Handwritten digits used for PCA training…

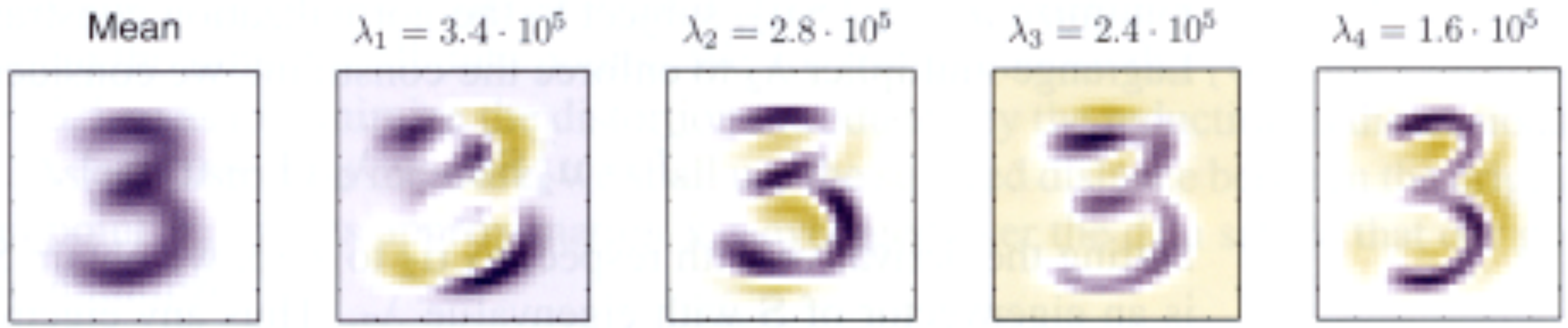
PCA



PRML - C. Bishop (Sec. 12.1)

PCA - Reconstruction

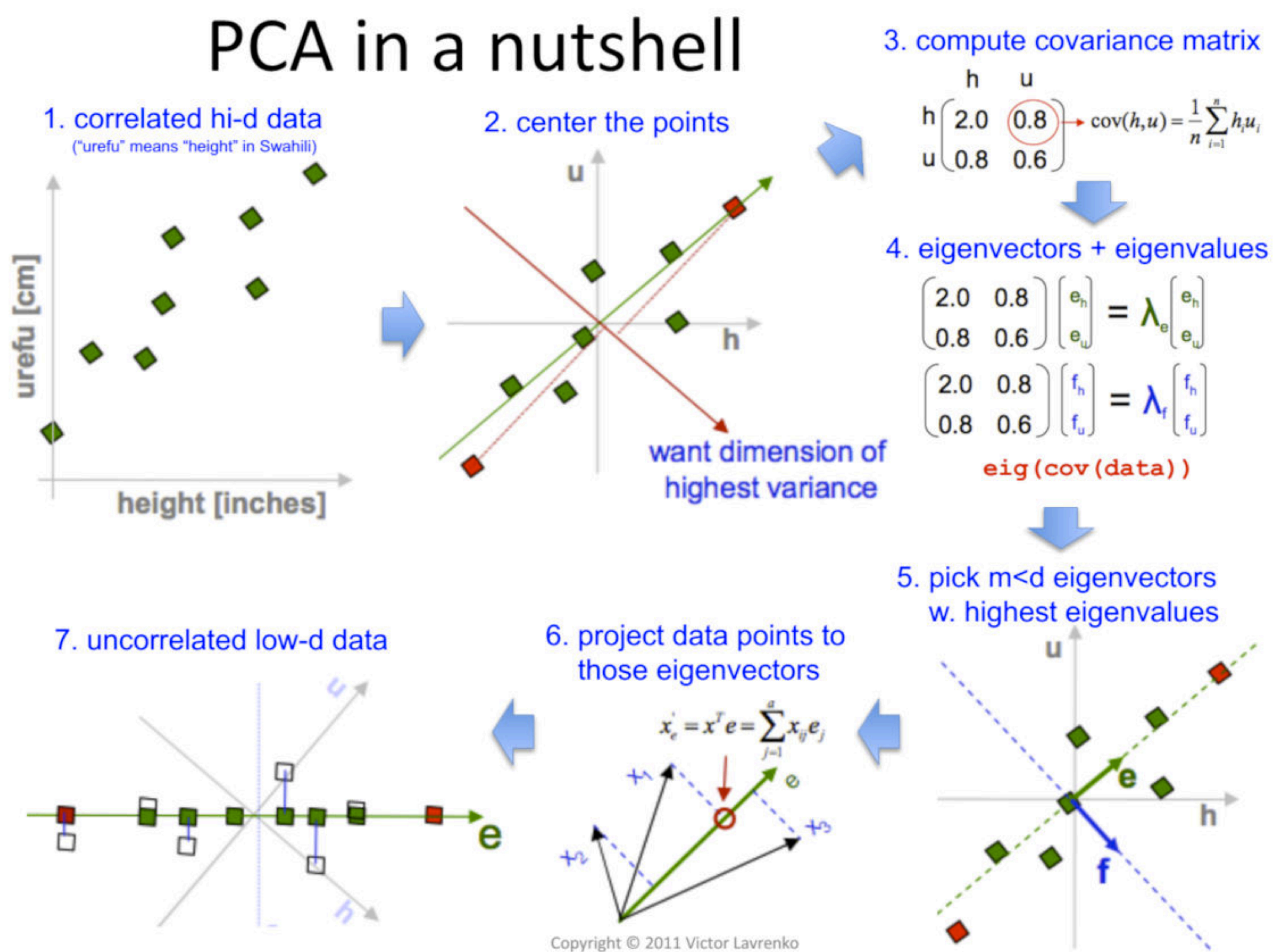
Eigenvectors



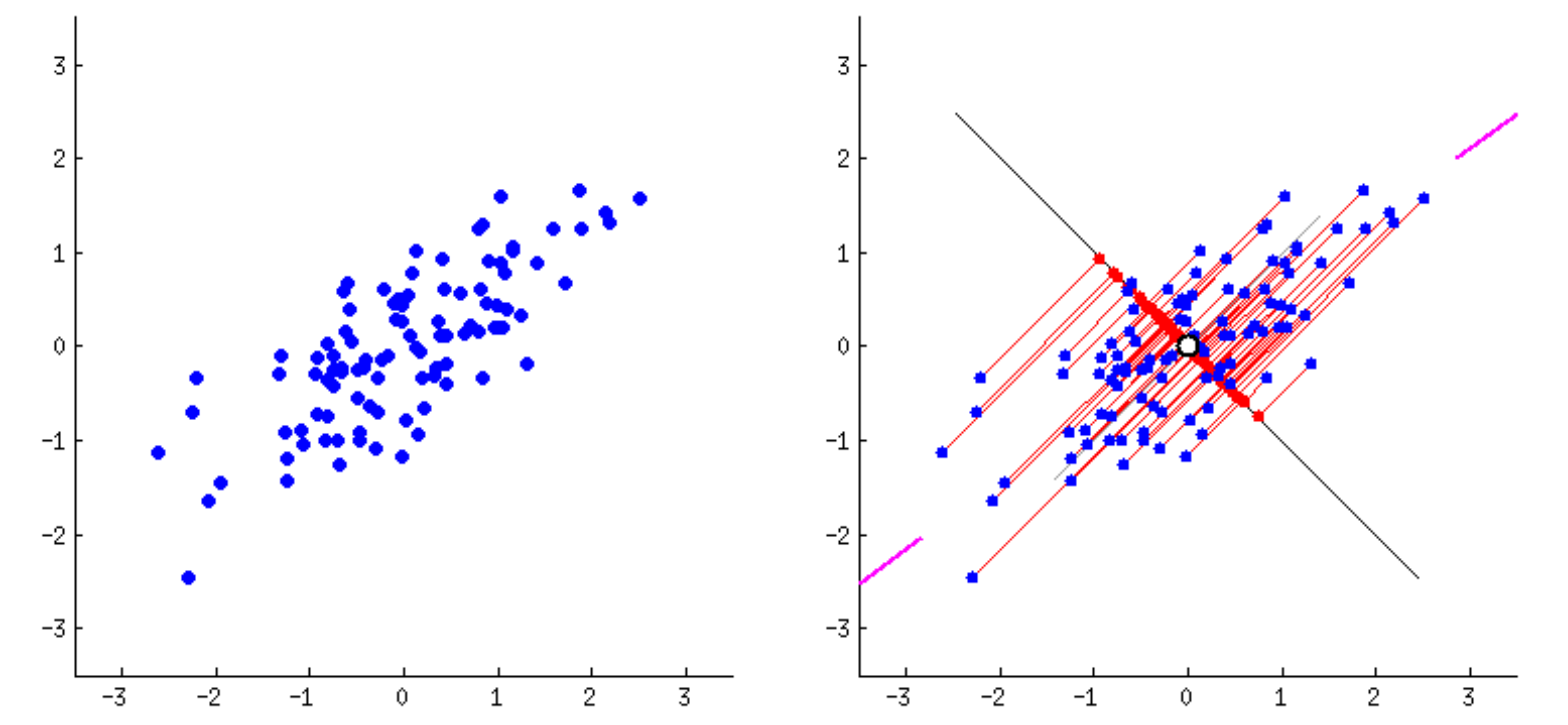
PCA - Reconstruction

PRML - C. Bishop (Sec. 12.1)

PCA Summary

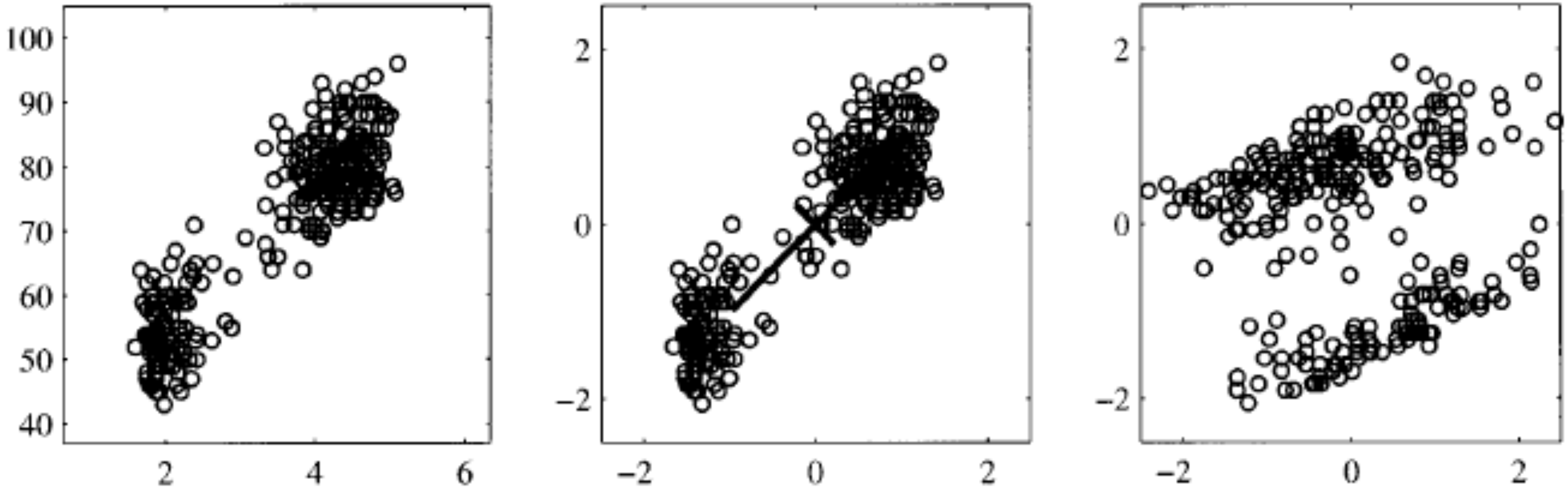

PCA in a nutshell
1. correlated hi-d data
("urefu" means "height" in Swahili)
urefu [cm]
height [inches]
2. center the points
u
h
want dimension of highest variance
3. compute covariance matrix
h u
h 2.0 0.8
u 0.8 0.6
$cov(h,u) = \frac{1}{n}\sum_{i=1}^{n} h_i u_i$
4. eigenvectors + eigenvalues
eig(cov(data))
5. pick m<d eigenvectors w. highest eigenvalues
e
f
6. project data points to those eigenvectors
$x_e' = x^T e = \sum_{j=1}^{d} x_{ij} e_j$
7. uncorrelated low-d data
Copyright © 2011 Victor Lavrenko

Visualizing PCA



Whitening the Data

Original Data | Standardization | Whitening



PRML - C. Bishop (Sec. 12.1)

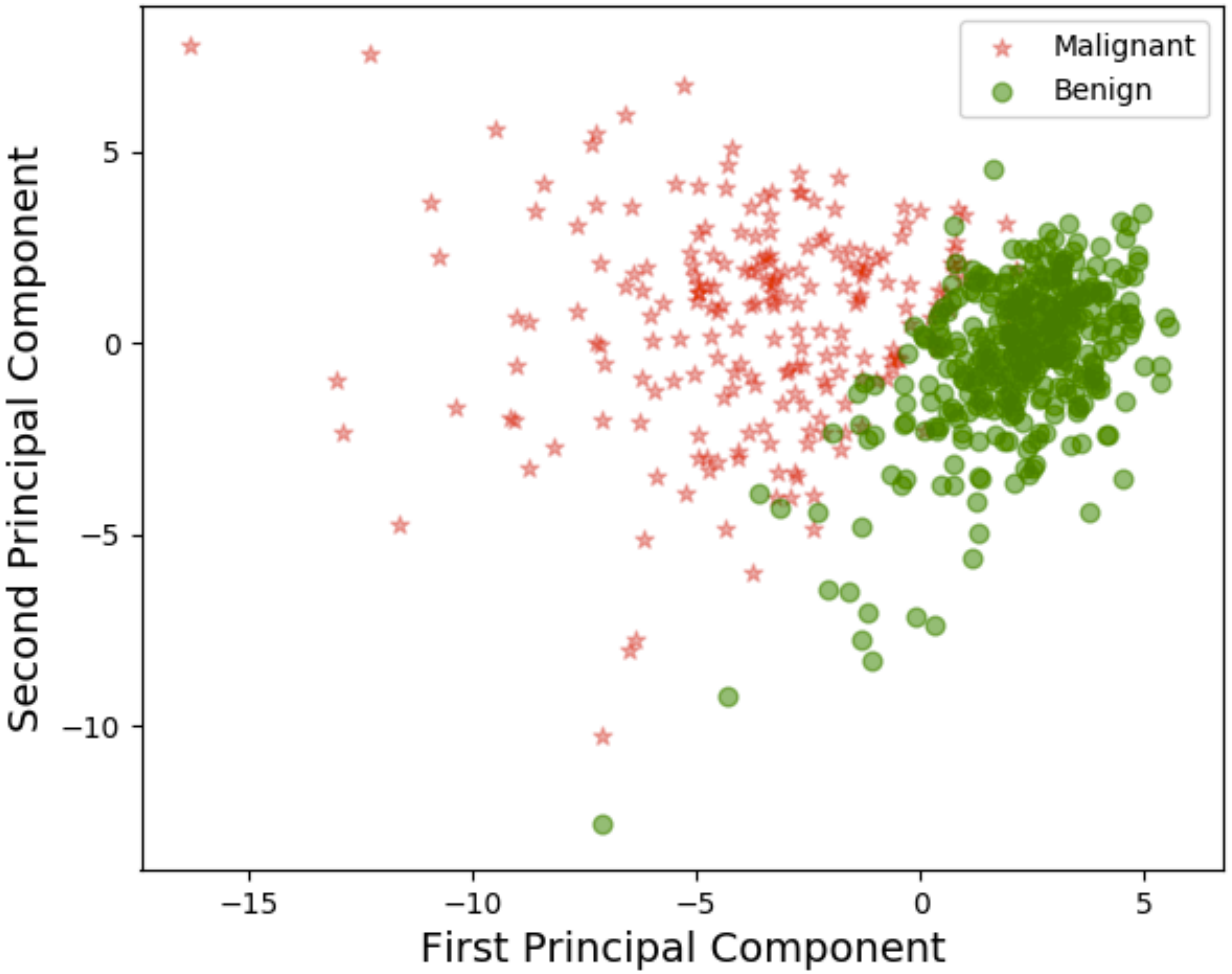
APPLICATION

- Wisconsin Cancer dataset (https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic))
- 569 participants
- 212 (M) 357 (B)
- 30 features —> digitized image of a fine needle aspirate (FNA) of a breast mass. The features describe characteristics of the cell nuclei present in the image.

Raw Features

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

PCA



Malignant
Benign
Second Principal Component
First Principal Component
5
0
−5
−10
−15
−10
−5
0
5

THANK YOU

Sriram Ganapathy and TA team
LEAP lab, C328, EE, IISc
sriramg@iisc.ac.in