E9 205 Machine Learning for Signal Processing

## Dimensionality Reduction - I

21-08-2019

Instructor - Sriram Ganapathy (sriramg@iisc.ac.in)

Principal Component Analysis

* Reducing the data $\mathbf{x}_n$ of dimension D to lower dimension $M < D$

* Projecting the data into subspace which preserves maximum data variance

  * Maximize variance in projected space

* Equivalent formulated as minimizing the error between the original and projected data points.

Direction of Maximum Variance
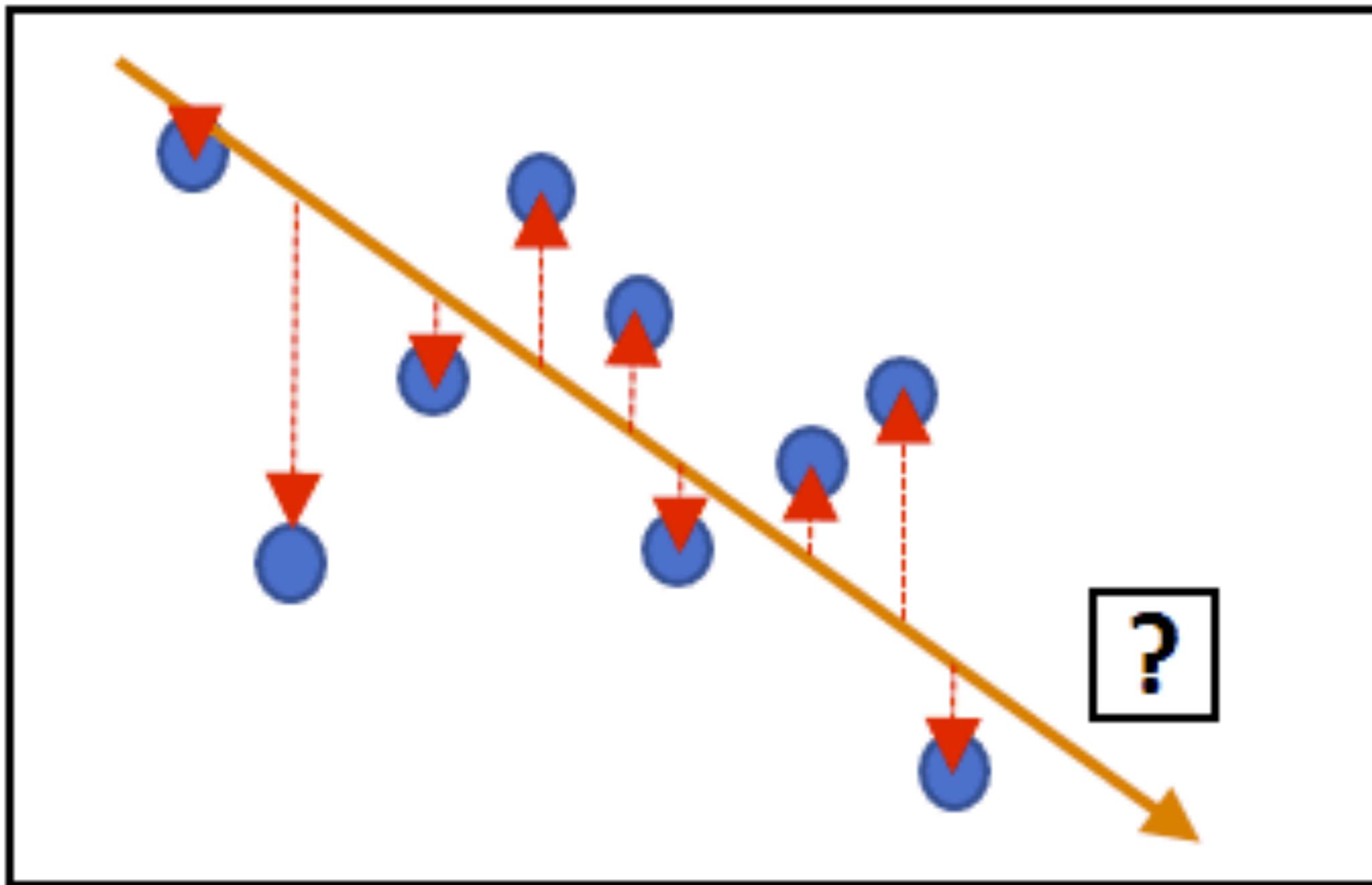
Minimum Error Formulation

PRML - C. Bishop (Sec. 12.1)
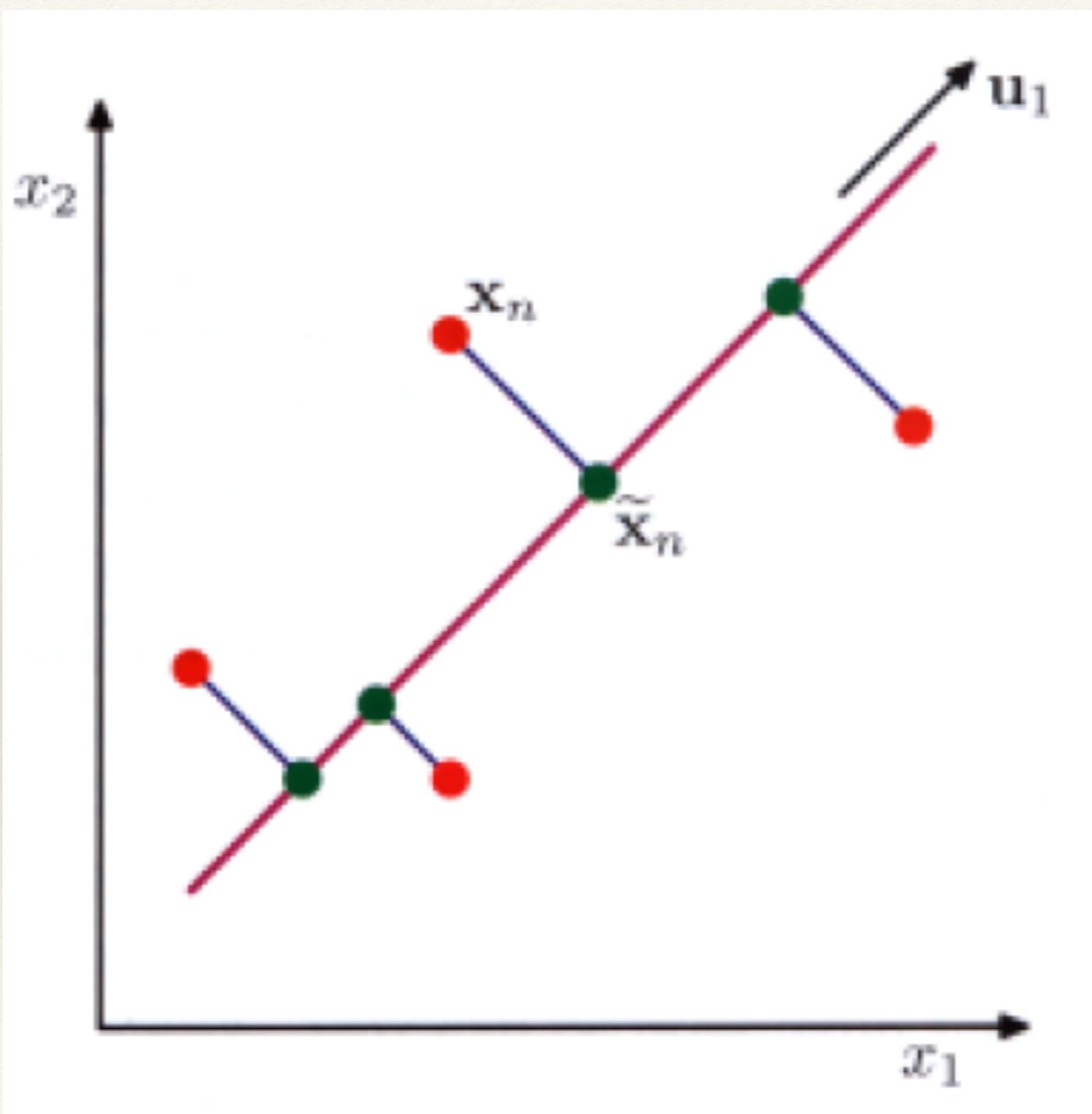
PCA Example
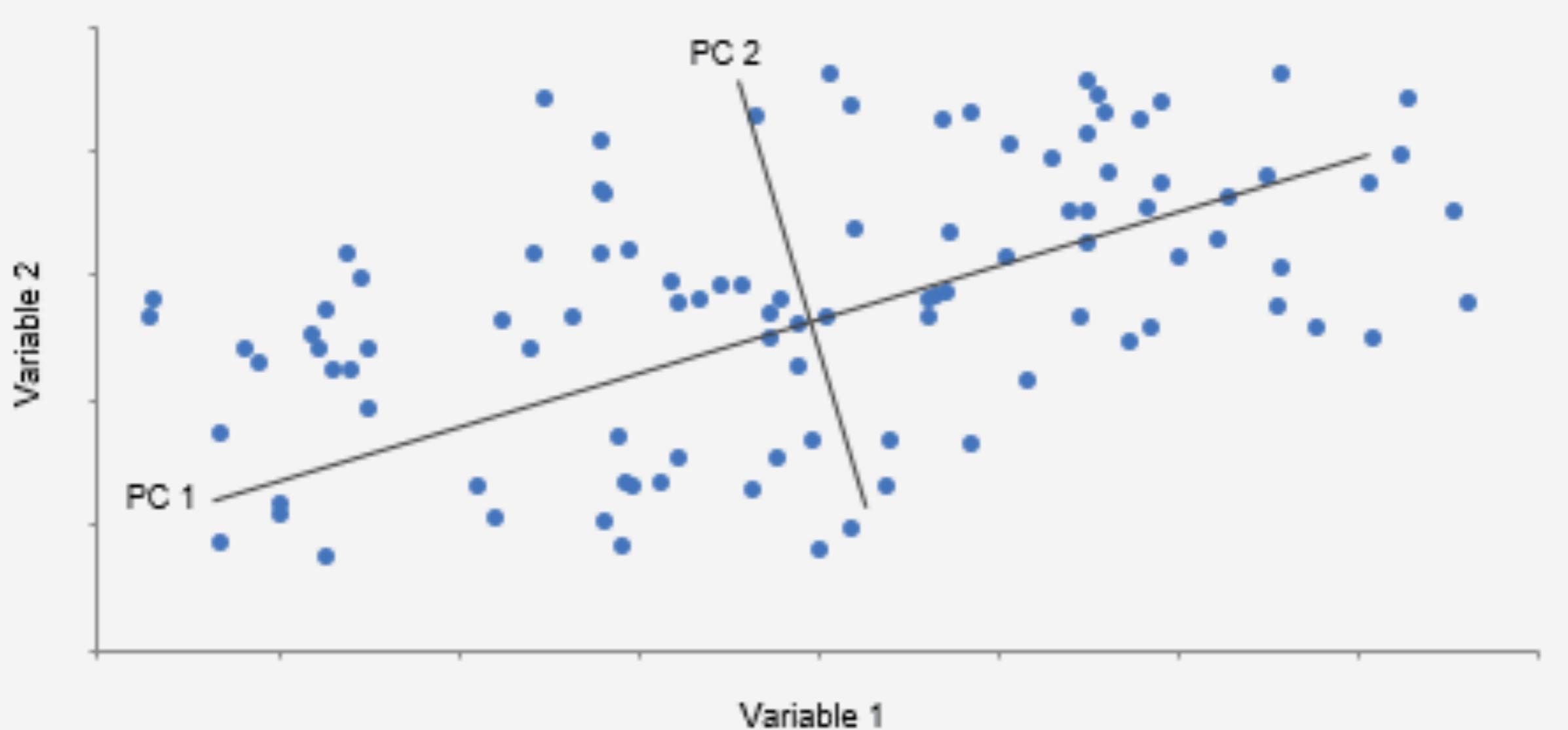
Principal Component Analysis

❖ 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

Handwritten digits used for PCA training…

PCA

PRML - C. Bishop (Sec. 12.1)
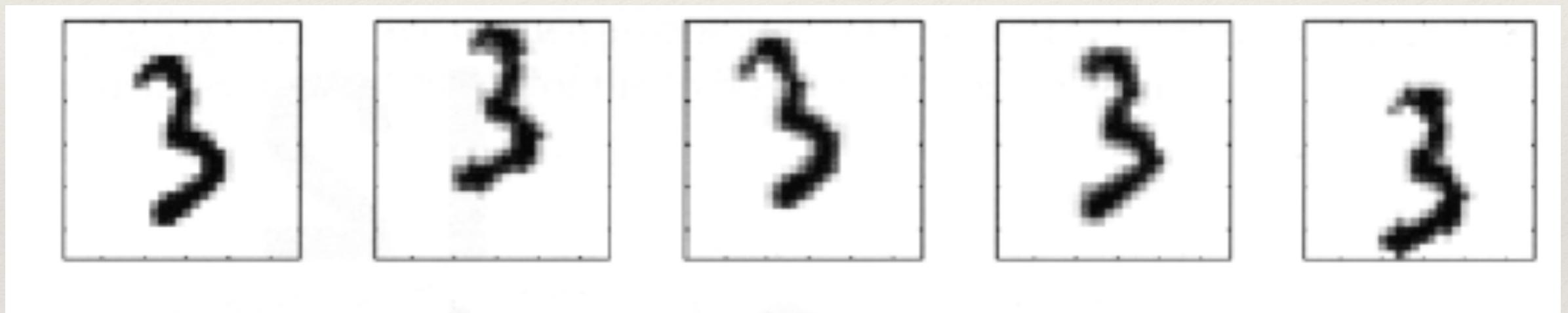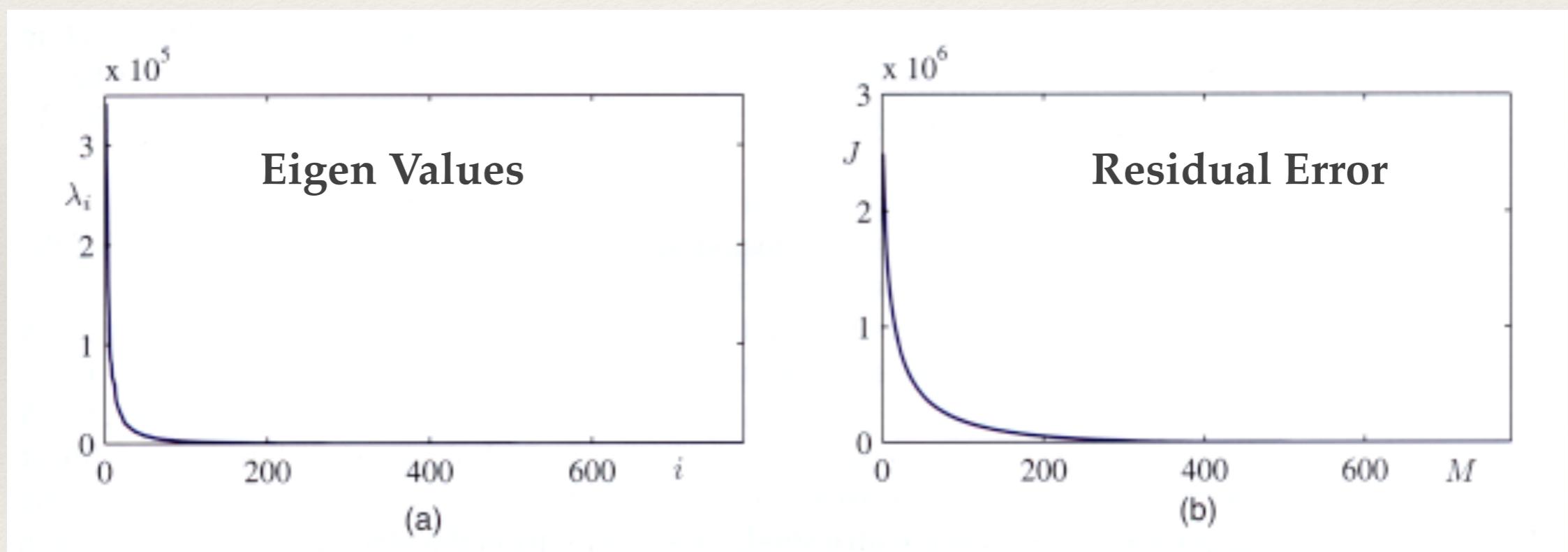
PCA - Reconstruction

Eigenvectors

| Mean | $\lambda_1 = 3.4 \cdot 10^5$ | $\lambda_2 = 2.8 \cdot 10^5$ | $\lambda_3 = 2.4 \cdot 10^5$ | $\lambda_4 = 1.6 \cdot 10^5$ |

PCA - Reconstruction

| Original | $M = 1$ | $M = 10$ | $M = 50$ | $M = 250$ |

PRML - C. Bishop (Sec. 12.1)
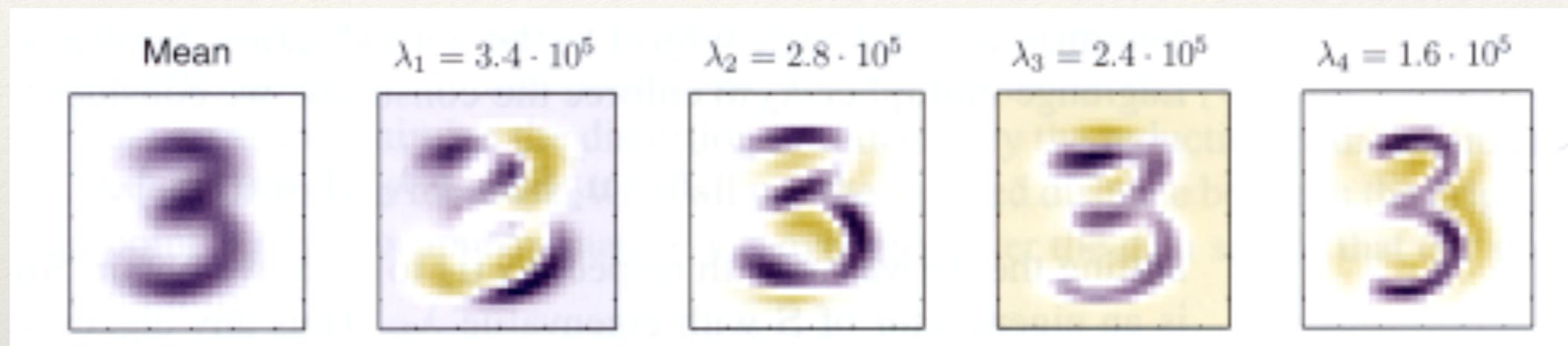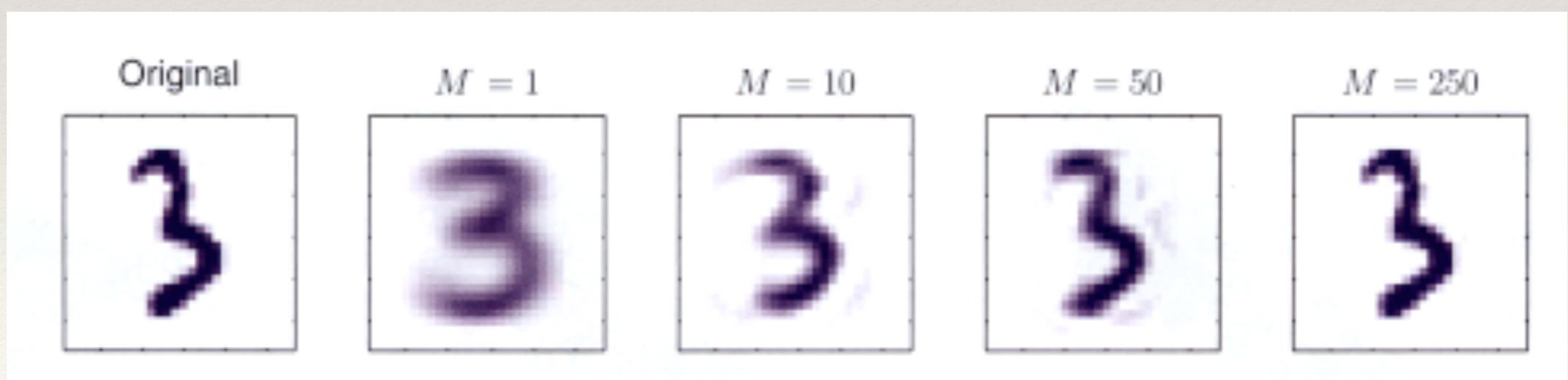
Whitening the Data
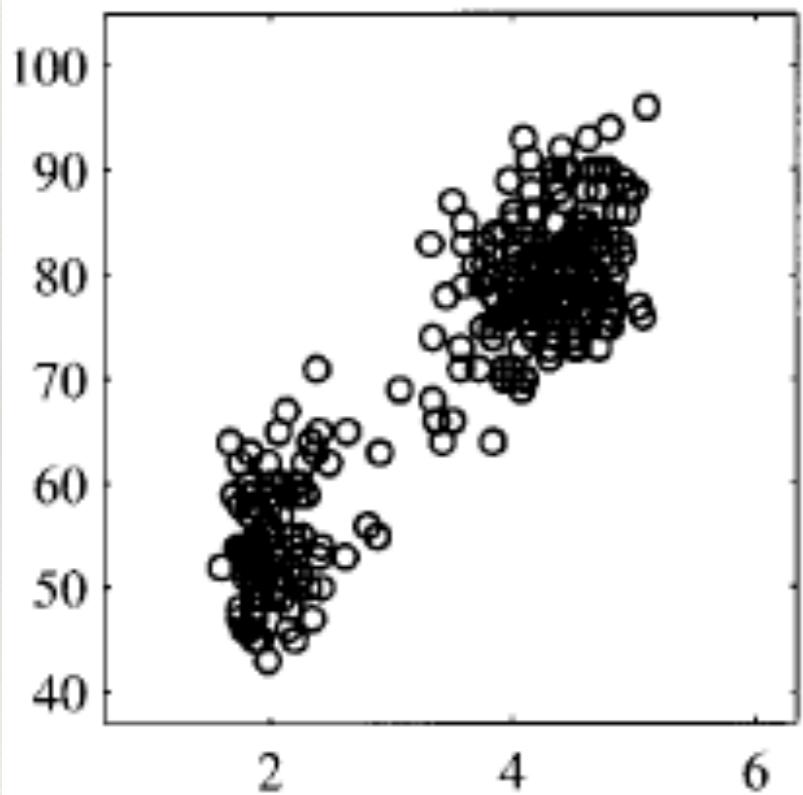
Original Data **Standardization** **Whitening**

PRML - C. Bishop (Sec. 12.1)