E9 205 Machine Learning for Signal Processing

Dimensionality Reduction - I

24-08-2017

Instructor - Sriram Ganapathy (sriram@ee.iisc.ernet.in)

Teaching Assistant - Aravind Illa (aravindece77@gmail.com).

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.

Minimum Error Formulation - PCA

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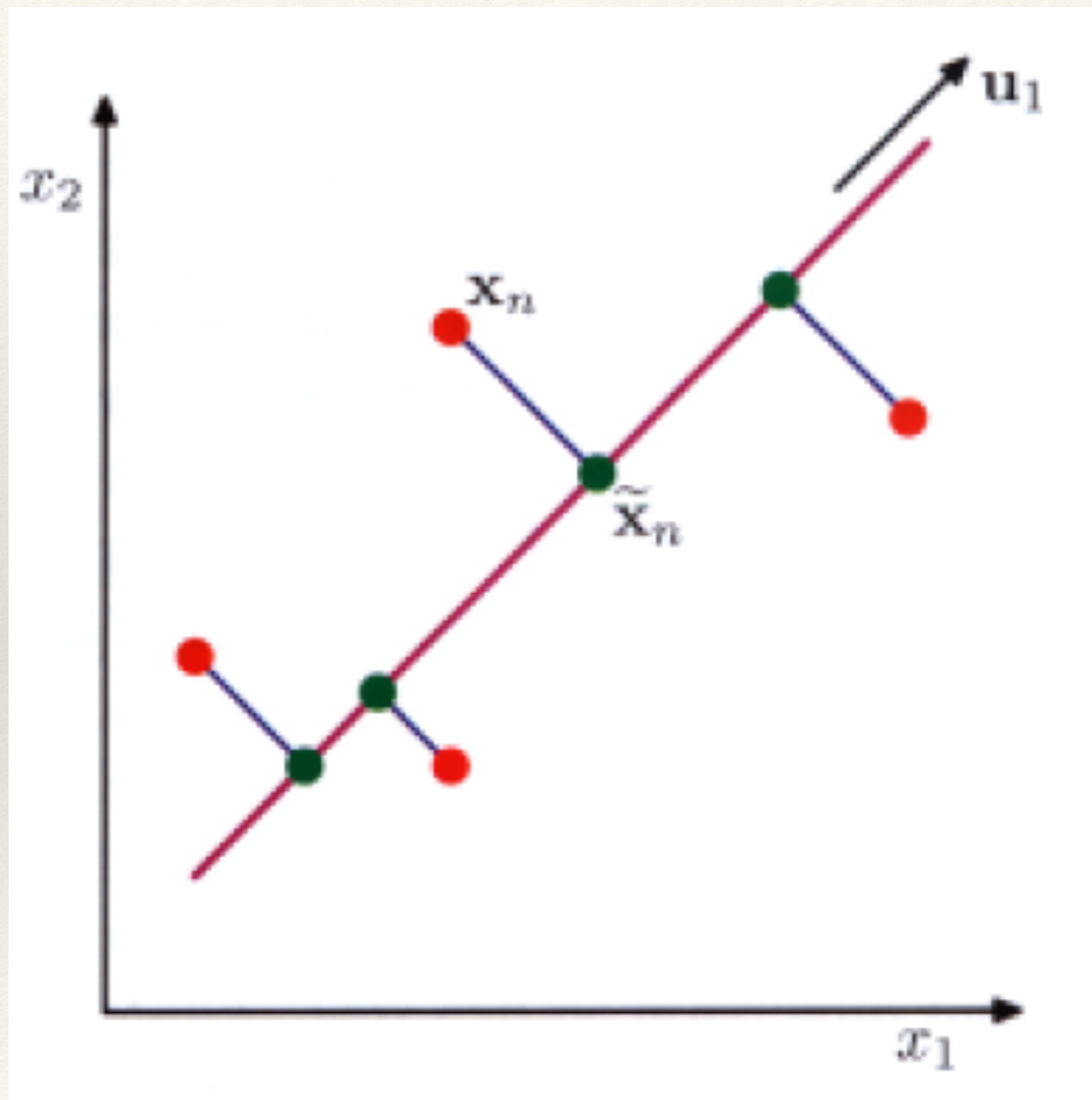
Principal Component Analysis

- First M eigenvectors of data covariance matrix

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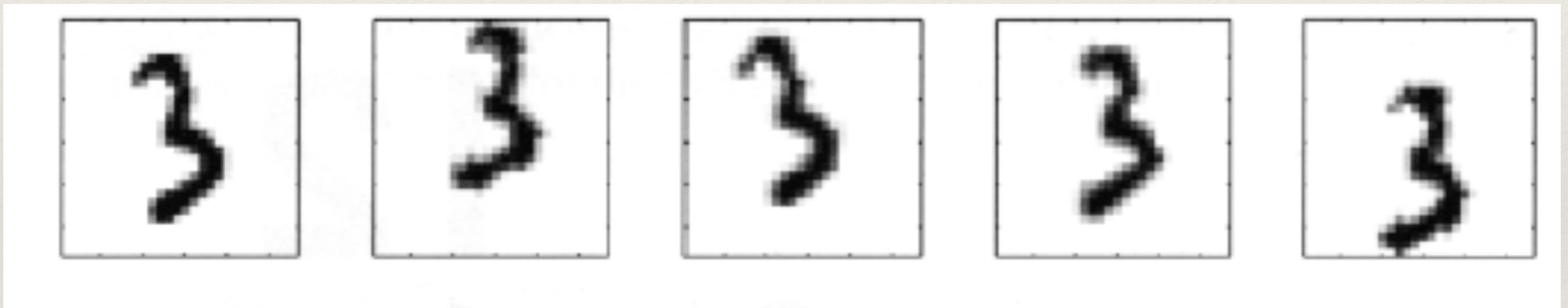
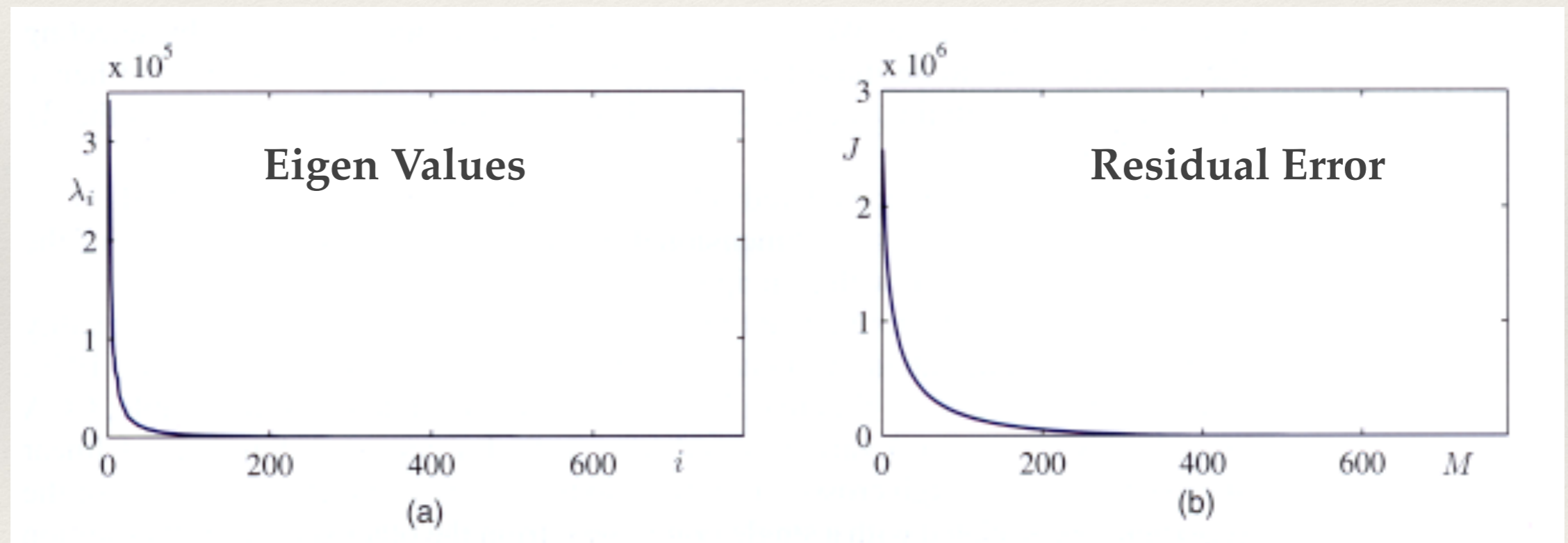- Residual error from PCA

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

PRML - C. Bishop (Sec. 12.1)

PCA

Handwritten digits used for PCA training…

PCA

PRML - C. Bishop (Sec. 12.1)

PCA - Reconstruction

Eigenvectors

PCA - Reconstruction

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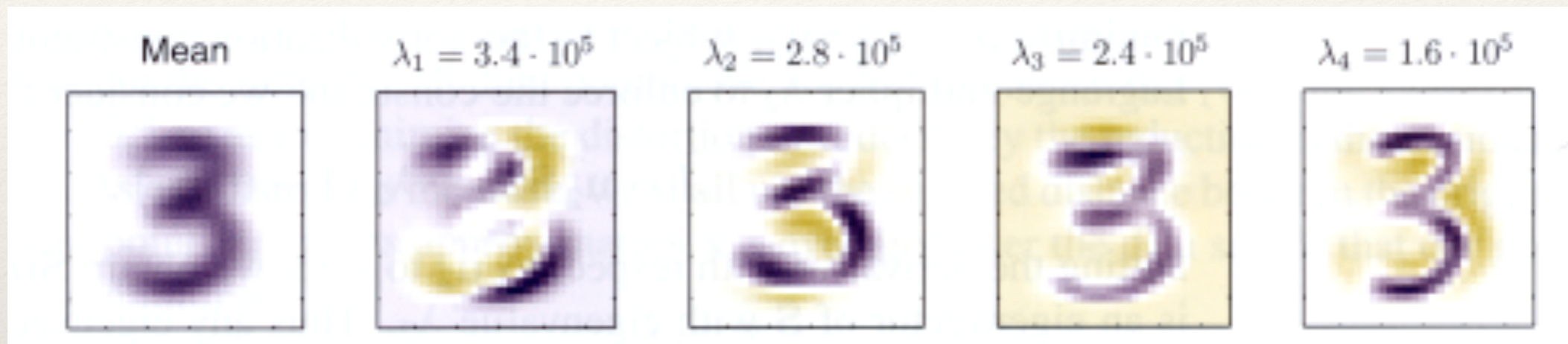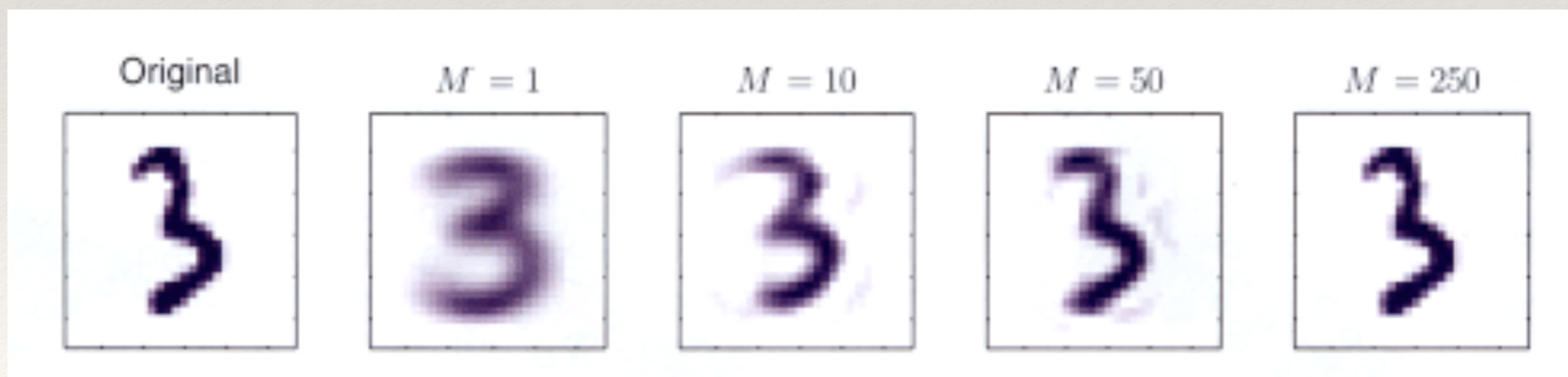
Whitening the Data
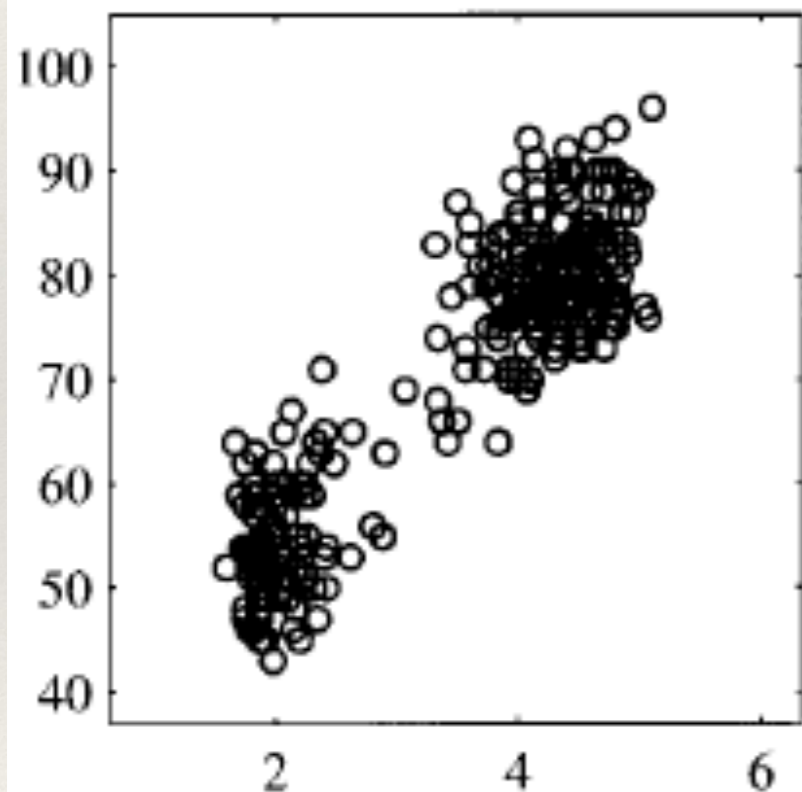
Original Data **Whitened data** **Whitened data**

PRML - C. Bishop (Sec. 12.1)