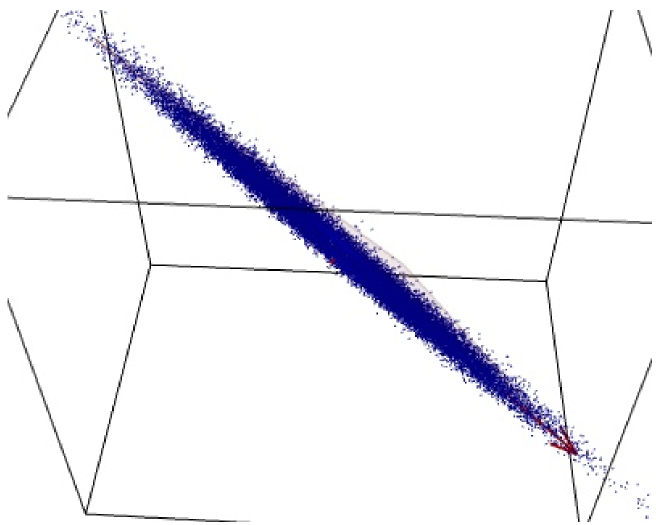
Lecture 32
Principal Component Analysis
ECEN 5283 Computer Vision

Dr. Guoliang Fan
School of Electrical and Computer Engineering
Oklahoma State University

Goals

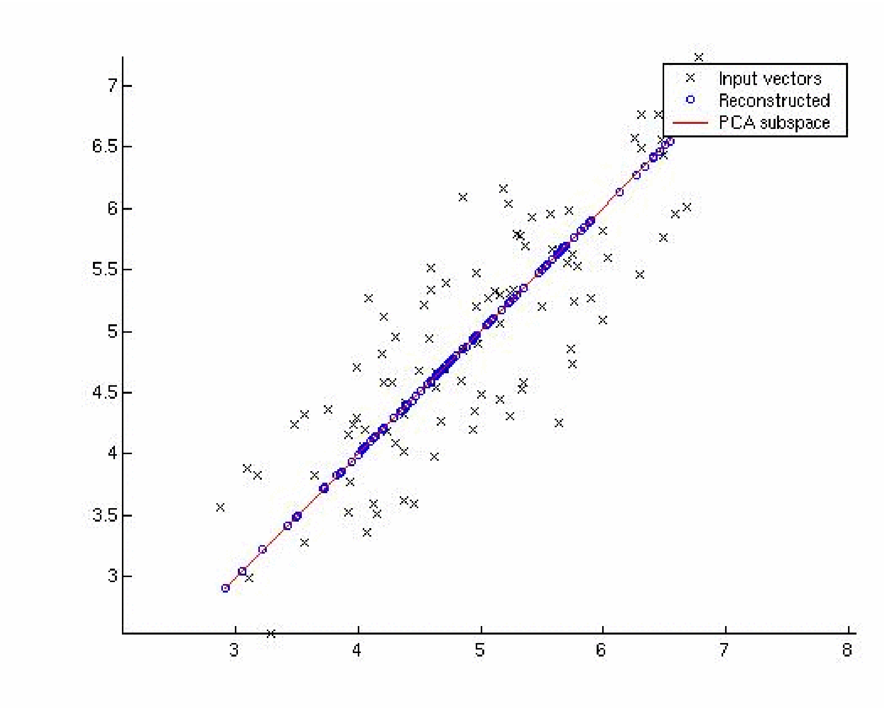
- To reduce the dimension of the data set for high-level vision tasks by linear Principal Component Analysis (PCA).

- To introduce some PCA-based applications.



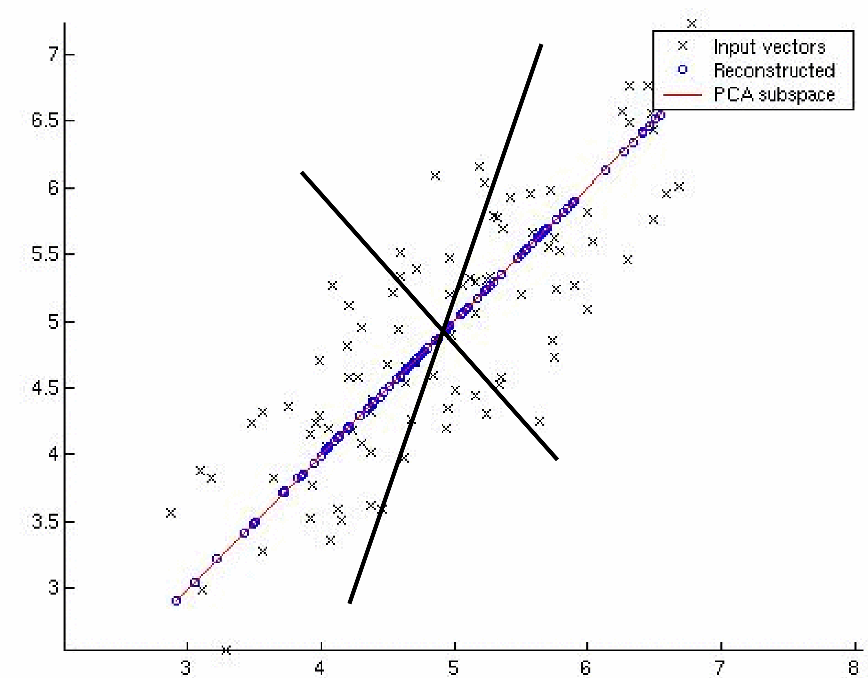
Principal Component Analysis (PCA)

- **PCA** *provides compact data representation*
  - We can construct a lower dimensional linear subspace that "best explains" the variation of these data points from their mean.
  - All data will be represented in this low-dimension feature space.



PCA: Optimality

- **PCA** *seeks the optimal projection that best represents the data* in a least-square sense for feature space *dimension reduction*.



PCA: Formulation

- A set of *n* feature vectors $x_i(i=1,\ldots,n)$ in $\Re^{\mathrm{d}}$.
  - The **mean** of this set of feature vectors is μ (center of gravity).
  - We use μ as an origin, and study the offsets from means,
$$(x_i - \mu).$$
  - The **covariance** of this set of feature vectors is $\sum$ defined as
$$\Sigma = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)(x_i - \mu)^T$$
- A unit feature vector **v** represents a direction in the original feature space.
  - The projection of a data point x_i on **v** is represented by
$$\boldsymbol{v} = \mathbf{v}(\boldsymbol{x_i}) = \mathbf{v}^{\boldsymbol{T}}(\mathbf{x_i} - \mu).$$
  - Notice that *v* has zero mean, and *v* should be able to capture the variance of the original data set.

PCA: Formulation (Cont'd)

- The variance of *v* is

$$\mathrm{var}(\boldsymbol{v}) = \frac{1}{\boldsymbol{n}-1}\sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \mathbf{v}(\boldsymbol{x}_{\boldsymbol{i}})\mathbf{v}(\boldsymbol{x}_{\boldsymbol{i}})^{\boldsymbol{T}}$$

$$= \frac{1}{\boldsymbol{n}-1}\sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \mathbf{v}^{\boldsymbol{T}}(\boldsymbol{x}_{\boldsymbol{i}}-\mu)(\mathbf{v}^{\boldsymbol{T}}(\boldsymbol{x}_{\boldsymbol{i}}-\mu))^{\boldsymbol{T}}$$

$$= \mathbf{v}^{T}\left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^{T}\right)\mathbf{v}$$

$$= \mathbf{v}^{\boldsymbol{T}}\sum\mathbf{v}$$

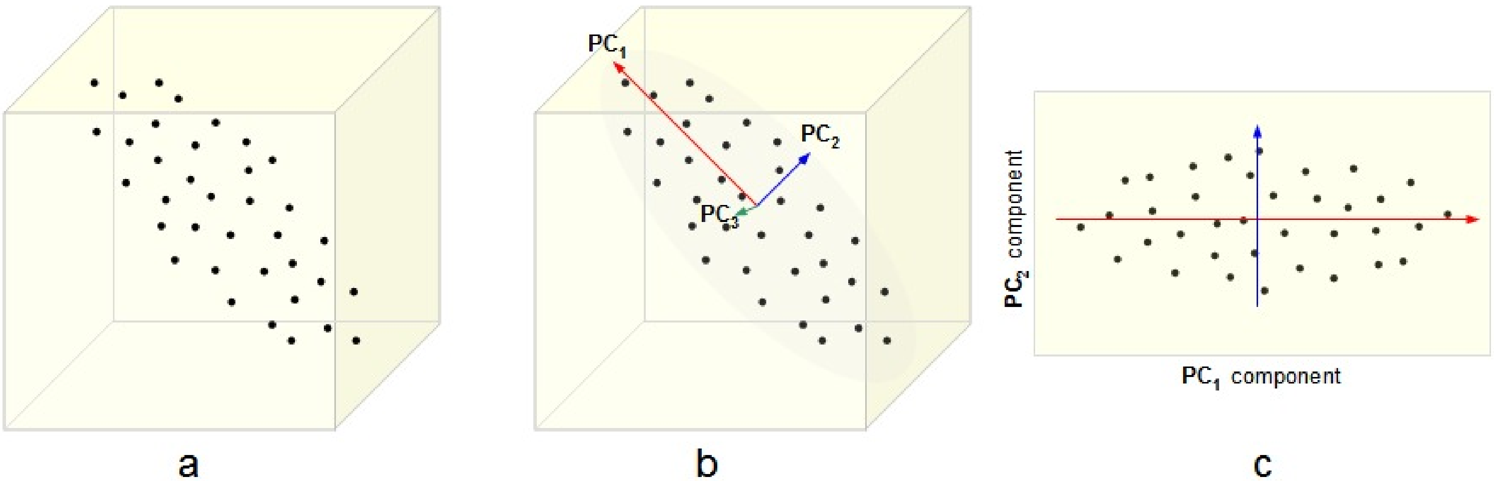
PCA: Solution

- We would like to maximize $\mathbf{v}^T \sum \mathbf{v}$ subject to $\mathbf{v}^T \mathbf{v} = 1$.

$$\Sigma = \mathbf{Q}\begin{bmatrix} \lambda_1^2 & 0 & 0 & 0 \\ 0 & \lambda_2^2 & 0 & 0 \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \lambda_q^2 \end{bmatrix}\mathbf{Q}^T \quad \text{with } 0 \le \lambda_1^2 \le \cdots \le \lambda_q^2 \text{ and } \mathbf{Q} = \begin{pmatrix} \mathbf{e}_1^T \\ \vdots \\ \mathbf{e}_q^T \end{pmatrix}$$

- This is an eigenvalue problem, and the eigenvector corresponding to the largest eigenvalue $\mathbf{e}_q$ is the solution.
- The eigenvectors associated large eigenvalues reveals the underlying data distribution.
- The accuracy of PCA is determined by the ratio between the sum of top largest eigenvalues and that of all eigenvalues.

PCA: Result



$$\Sigma = \mathbf{Q}\begin{pmatrix} \lambda_1^2 & 0 & 0 \\ 0 & \lambda_2^2 & 0 \\ 0 & 0 & \lambda_3^2 \end{pmatrix}\mathbf{Q}^T \quad \text{with } \lambda_1^2 \geq \lambda_2^2 \geq \lambda_3^2 \text{(ascendent order) and } \mathbf{Q} = \begin{pmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \\ \mathbf{e}_3^T \end{pmatrix}$$

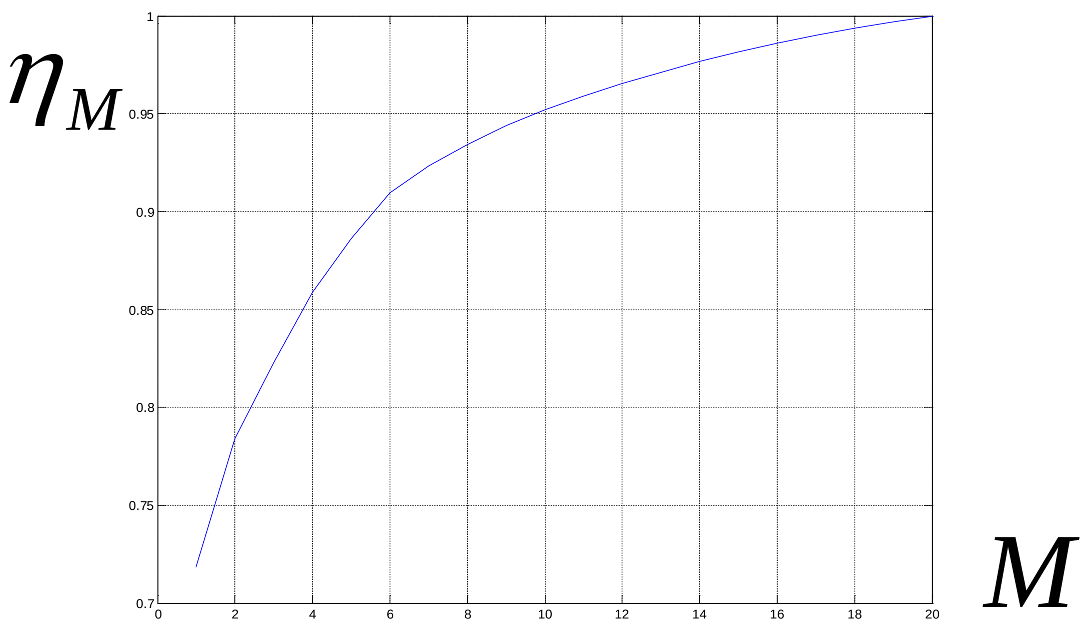
If someone ask you what the lowest dimension needed?

You should ask him how much data variation (energy) to be preserved?

M-Dimensional Subspace

- For an M-dimensional projection space, the optimal linear projection is defined by the *top M eigenvectors* associated with the M largest eigenvalues.

$$\eta_M = \frac{J_M}{J_D} = \frac{\sum_{i=1}^{M} \lambda_i}{\sum_{j=1}^{D} \lambda_j}$$



PCA: Examples

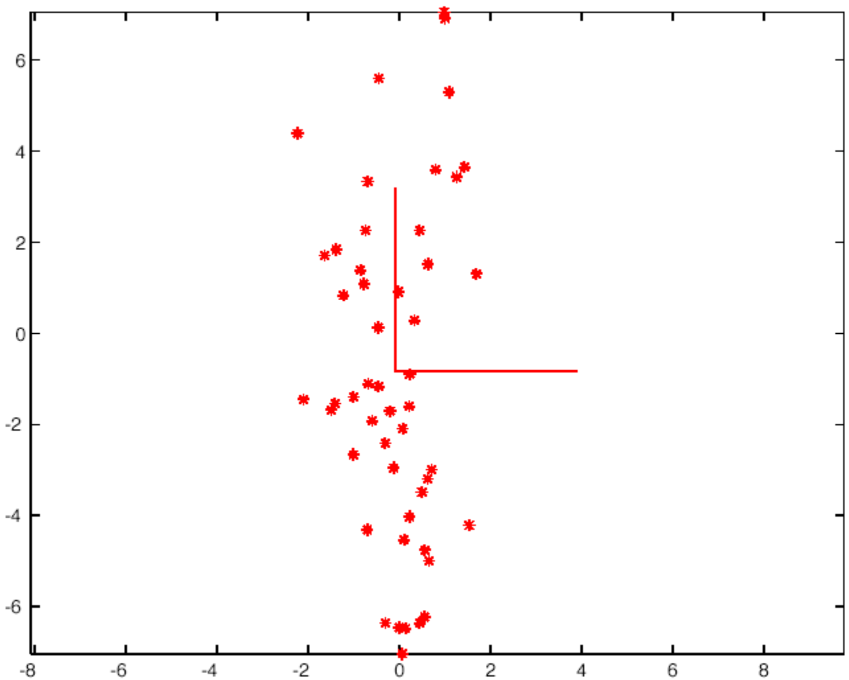


Figure 25.6. A data set which is well represented by a principal component analysis. The axes represent the directions obtained using PCA; the vertical axis is the first principal component, and is the direction in which the variance is highest.

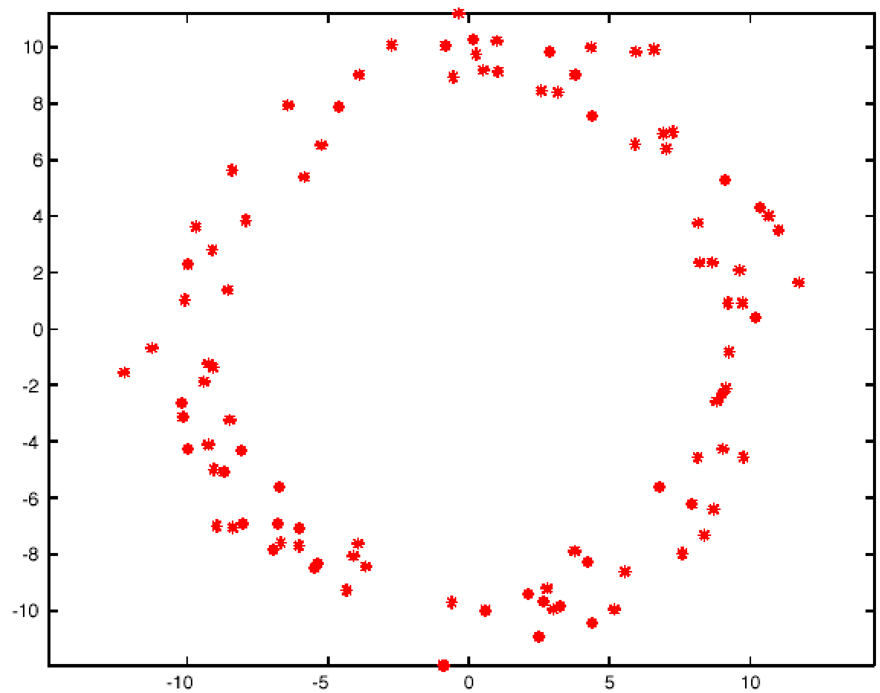


Figure 25.7. Not every data set that is well represented by PCA. The principal components of this data set will be relatively unstable, because the variance in each direction is the same for the source. This means that we may well report significantly different principal components for different datasets from this source. This is a secondary issue — the main difficulty is that projecting the data set onto some axis will suppress the main feature, its circular structure.

PCA Applications: Data Compression (1)

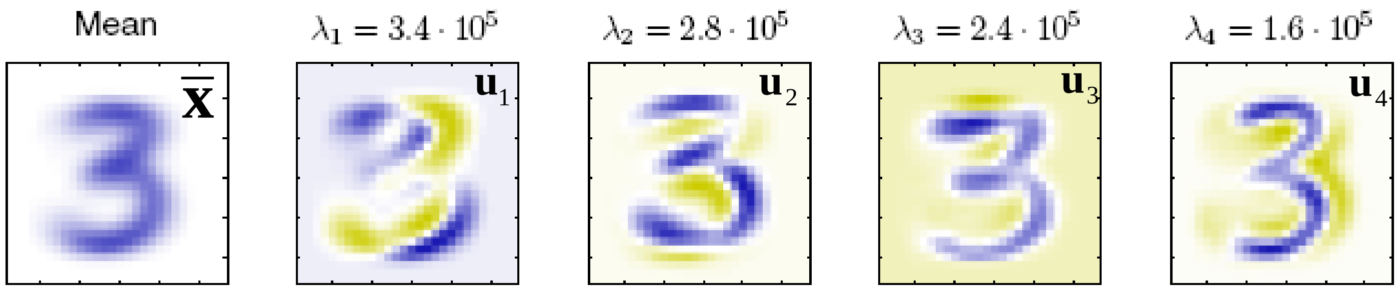


Figure 12.3 The mean vector $\overline{\mathbf{x}}$ along with the first four PCA eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_4$ for the off-line digits data set, together with the corresponding eigenvalues.

$$\mathbf{S} = \frac{1}{N}\left(\sum_{n=1}^{N}(\mathbf{x}_n - \overline{\mathbf{x}})(\mathbf{x}_n - \overline{\mathbf{x}})^{\mathrm{T}}\right) \quad \text{where } \overline{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n$$

$$\mathbf{S} = \mathbf{U}^{\mathbf{T}}\mathbf{L}\mathbf{U} \qquad \mathbf{L} = \mathrm{Diag}(\lambda_1, \lambda_2, \ldots, \lambda_D) \qquad (\lambda_1 > \lambda_2 > \ldots > \lambda_D)$$

$$\mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}} = \mathbf{L}$$

(eigenvalue analysis) $\mathbf{U} = \begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_D\end{pmatrix}$

PCA Applications: Data Compression (2)

$$\mathbf{u}_i^{\mathrm{T}}\mathbf{S}\mathbf{u}_i = \lambda_i \qquad \hat{J}_M = \sum_{i=M+1}^{D} \lambda_i$$

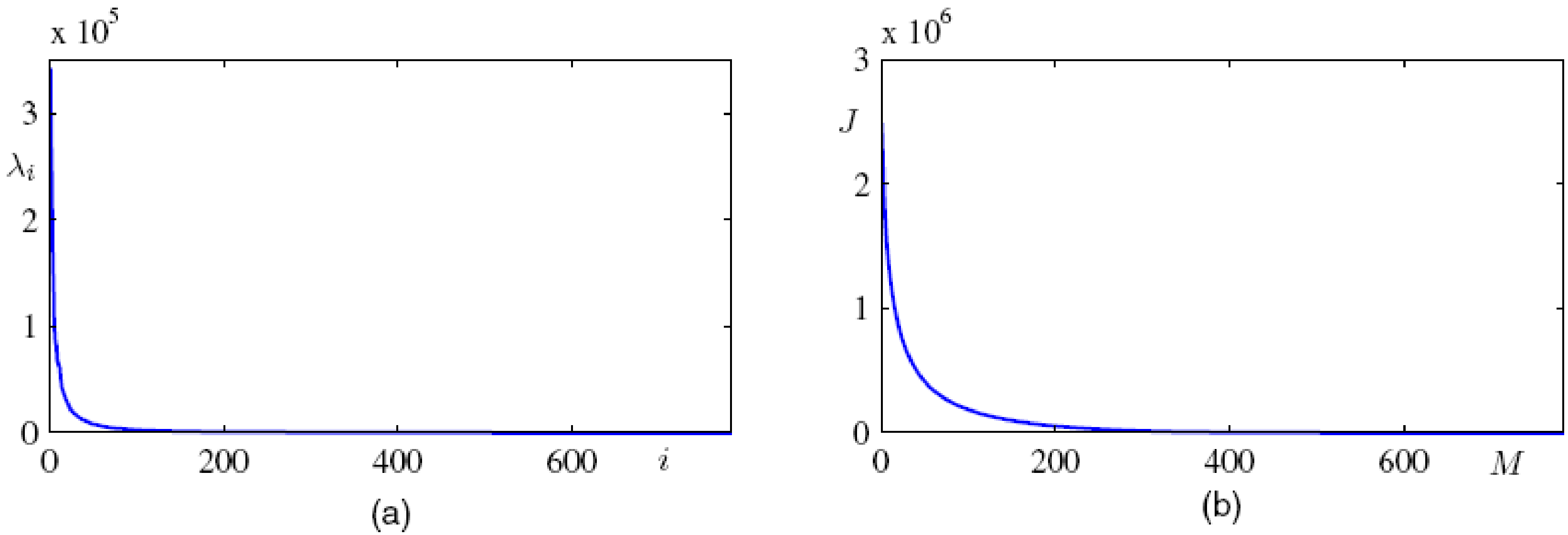


Figure 12.4 (a) Plot of the eigenvalue spectrum for the off-line digits data set. (b) Plot of the sum of the discarded eigenvalues, which represents the sum-of-squares distortion J introduced by projecting the data onto a principal component subspace of dimensionality M.

PCA Applications: Data Compression (3)

$$\hat{\mathbf{x}}_M = \bar{\mathbf{x}} + \sum_{m=1}^{M} \left\langle (\mathbf{x} - \bar{\mathbf{x}}), \mathbf{u}_m \right\rangle \mathbf{u}_m \quad \text{(PCA reconstruction)}$$

$$= \bar{\mathbf{x}} + \sum_{m=1}^{M} \left((\mathbf{x} - \bar{\mathbf{x}})^{\mathrm{T}} \mathbf{u}_m \right) \mathbf{u}_m$$

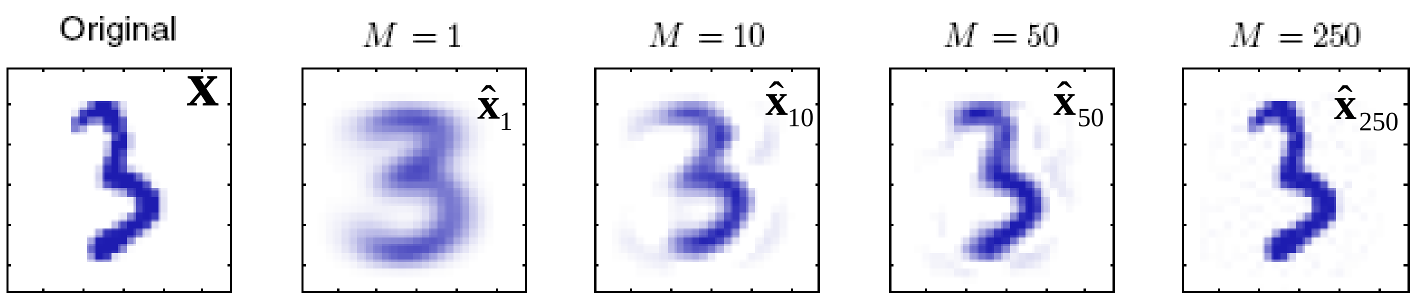
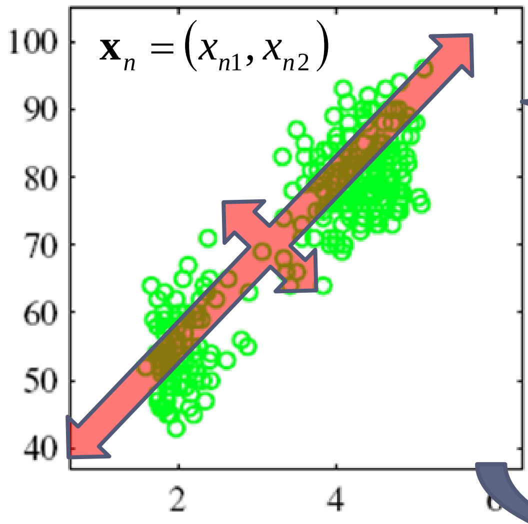


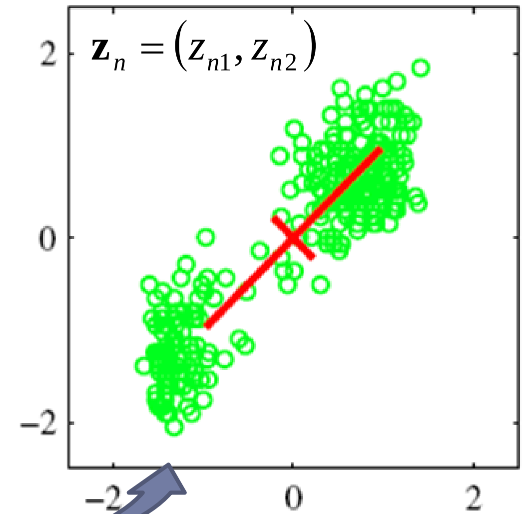
Figure 12.5 An original example from the off-line digits data set together with its PCA reconstructions obtained by retaining M principal components for various values of M. As M increases the reconstruction becomes more accurate and would become perfect when $M = D = 28 \times 28 = 784$.

PCA Applications: Data Pre-processing (1)



Problem 1:
Large difference on the range of two variables

Problem 2:
Strong correlation between the two variables



Normalization

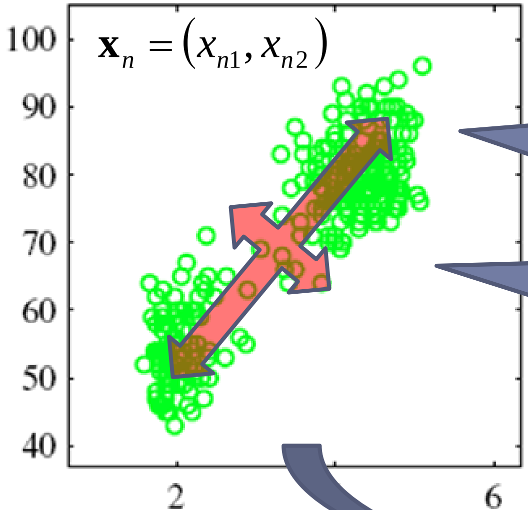
$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n$$

$$m_i = \frac{1}{N}\sum_{n=1}^{N} x_{ni} \qquad v_i = \frac{1}{N}\sum_{n=1}^{N}(x_{ni} - m_i)^2$$

$$z_{ni} = \frac{x_{ni} - m_i}{\sqrt{v_i}}$$

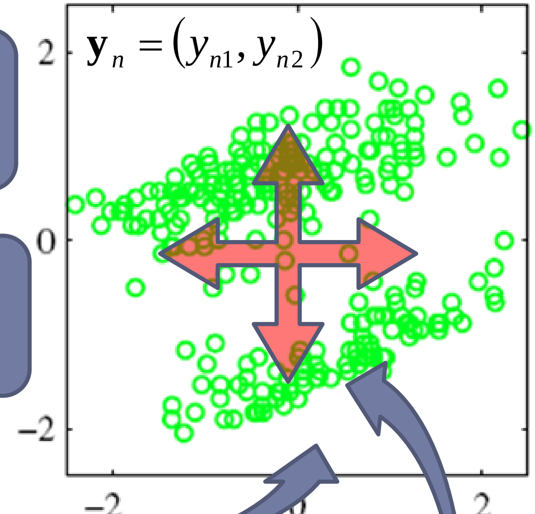
$$\mathbf{z}_n = (z_{n1}, z_{n2})$$

PCA Applications: Data Pre-processing (2)



Problem 1:
Large difference on the range of two variables

Problem 2:
Strong correlation between the two variables



Whitening

$$\mathbf{y}_n = \mathbf{L}^{-1/2}\mathbf{U}^{\mathrm{T}}(\mathbf{x}_n - \overline{\mathbf{x}})$$

$$\mathbf{L} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

$$\mathbf{U} = (\mathbf{u}_1 \quad \mathbf{u}_2)$$

$$\mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}} = \mathbf{L}$$

$$\frac{1}{N}\sum_{n=1}^{N}\mathbf{y}_n\mathbf{y}_n^{\mathrm{T}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{L}^{-1/2}\mathbf{U}^{\mathrm{T}}(\mathbf{x}_n - \overline{\mathbf{x}})(\mathbf{x}_n - \overline{\mathbf{x}})^{\mathrm{T}}\mathbf{U}\mathbf{L}^{-1/2}$$

$$= \frac{1}{N}\mathbf{L}^{-1/2}\mathbf{U}^{\mathrm{T}}\left(\sum_{n=1}^{N}(\mathbf{x}_n - \overline{\mathbf{x}})(\mathbf{x}_n - \overline{\mathbf{x}})^{\mathrm{T}}\right)\mathbf{U}\mathbf{L}^{-1/2}$$

$$= \mathbf{L}^{-1/2}\boxed{\mathbf{U}^{\mathrm{T}}\mathbf{S}\mathbf{U}}\mathbf{L}^{-1/2} = \left(\mathbf{L}^{-1/2}\right)\mathbf{L}\left(\mathbf{L}^{-1/2}\right) = \mathbf{I}$$