



Machine Learning 1

Lecture 10.2 - Unsupervised Learning
Principal Component Analysis -
Reconstruction Error Minimization

Erik Bekkers

(Bishop 12.1.2, 12.1.3)



PCA via minimal reconstruction error

- Same method, alternative view to the maximal variance viewpoint

- PCA can be obtained by minimizing the reconstruction error. Find a transformation that minimizes:

$$\frac{1}{N} \sum_{n=1}^N \|\mathbf{x}_n - \tilde{\mathbf{x}}_n\|$$

$\tilde{\mathbf{x}}_n = \mathbf{U}_M \mathbf{z}_n + \bar{\mathbf{x}}$
 basis $\mathbb{R}^{p \times m}$
 latent variable / basis coeffs \mathbb{R}^m
 Before we focussed on embedding
 $\mathbf{z}_n = \mathbf{U}_M^T \mathbf{x}_n$

- Represent the points in a different orthonormal basis (unknown for now):

$$\{\mathbf{u}_i : \mathbf{u}_i^T \mathbf{u}_i = 1\}_{i=1}^D \quad \mathbf{u}_i^T \mathbf{u}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

PCA via minimal reconstruction error

‣ In the new basis $\mathbf{x}_n = \sum_{i=1}^D \alpha_{ni} \mathbf{u}_i$

property of
orthonormal
basis

‣ The coefficients are found by multiplying with \mathbf{u}_j^T on both sides:

$$\alpha_{ni} = \mathbf{x}_n^T \mathbf{u}_i$$

‣ Therefore... we get the projection onto the new basis:

$$\mathbf{x}_n = \sum_{i=1}^D \underbrace{(\mathbf{x}_n^T \mathbf{u}_i)}_{\alpha_{ni}} \mathbf{u}_i$$

‣ For reconstruction, we use the first M elements of the basis and a shared offset for the remaining dimensions

$$\tilde{\mathbf{x}}_n = \sum_{i=1}^M \underbrace{(\mathbf{x}_n^T \mathbf{u}_i)}_{z_{ni}} \mathbf{u}_i + \sum_{i=M+1}^D b_i \mathbf{u}_i$$

shared for all data points

PCA via minimal reconstruction error

- ▶ The difference/error is then given by *first m components cancel out*

$$\begin{aligned} \mathbf{x}_n - \tilde{\mathbf{x}}_n &= \underbrace{\sum_{i=1}^D (\mathbf{x}_n^T \mathbf{u}_i) \mathbf{u}_i}_{\tilde{\mathbf{x}}_n} - \underbrace{\sum_{i=1}^M (\mathbf{x}_n^T \mathbf{u}_i) \mathbf{u}_i}_{\tilde{\mathbf{x}}_n} - \sum_{i=M+1}^D b_i \mathbf{u}_i \\ &= \sum_{i=M+1}^D (\mathbf{x}_n^T \mathbf{u}_i) \mathbf{u}_i - \sum_{i=M+1}^D b_i \mathbf{u}_i \\ &= \sum_{i=M+1}^D (\mathbf{x}_n^T \mathbf{u}_i) \mathbf{u}_i - b_i \mathbf{u}_i \end{aligned}$$

PCA via minimal reconstruction error

- ▶ We have to minimize for both \mathbf{u}_i and b_i

$$\frac{1}{N} \sum_{n=1}^N \|\mathbf{x}_n - \tilde{\mathbf{x}}_n\|^2 = \frac{1}{N} \sum_{n=1}^N \left\| \sum_{i=M+1}^D (\mathbf{x}_n^T \mathbf{u}_i) \mathbf{u}_i - b_i \mathbf{u}_i \right\|^2$$

- ▶ Solution for b_i is given by $b_i = \bar{x}^T \mathbf{u}_i$
- ▶ We are left to solve for \mathbf{u}_i , minimizing

$$\frac{1}{N} \sum_{n=1}^N \left\| \sum_{i=M+1}^D ((\mathbf{x}_n - \tilde{\mathbf{x}})^T \mathbf{u}_i) \mathbf{u}_i \right\|^2$$

PCA via minimal reconstruction error

$$\begin{aligned}
 & \frac{1}{N} \sum_{n=1}^N \left\| \sum_{i=M+1}^D ((\mathbf{x}_n - \tilde{\mathbf{x}})^T \mathbf{u}_i) \mathbf{u}_i \right\|^2 \\
 & \stackrel{\text{expand square}}{=} \frac{1}{N} \sum_{n=1}^N \left(\sum_{i=M+1}^D \underbrace{((\mathbf{x}_n - \tilde{\mathbf{x}})^T \mathbf{u}_i)}_{\text{scalar}} \underbrace{\mathbf{u}_i}_{\text{vector}} \right)^T \left(\sum_{j=M+1}^D ((\mathbf{x}_n - \tilde{\mathbf{x}})^T \mathbf{u}_j) \mathbf{u}_j \right) \\
 & = \frac{1}{N} \sum_{n=1}^N \sum_{i=M+1}^D \sum_{j=M+1}^D \underbrace{((\mathbf{x}_n - \tilde{\mathbf{x}})^T \mathbf{u}_i)}_{\text{scalar}} \underbrace{\mathbf{u}_i^T \mathbf{u}_j}_{\substack{\text{0 for all } i \neq j \\ \text{1 for } i=j}} ((\mathbf{x}_n - \tilde{\mathbf{x}})^T \mathbf{u}_j) \\
 & = \frac{1}{N} \sum_{n=1}^N \sum_{i=M+1}^D \mathbf{u}_i^T (\mathbf{x}_n - \tilde{\mathbf{x}}) (\mathbf{x}_n - \tilde{\mathbf{x}})^T \mathbf{u}_i = \sum_{i=M+1}^D \mathbf{u}_i^T \mathbf{S} \mathbf{u}_i
 \end{aligned}$$

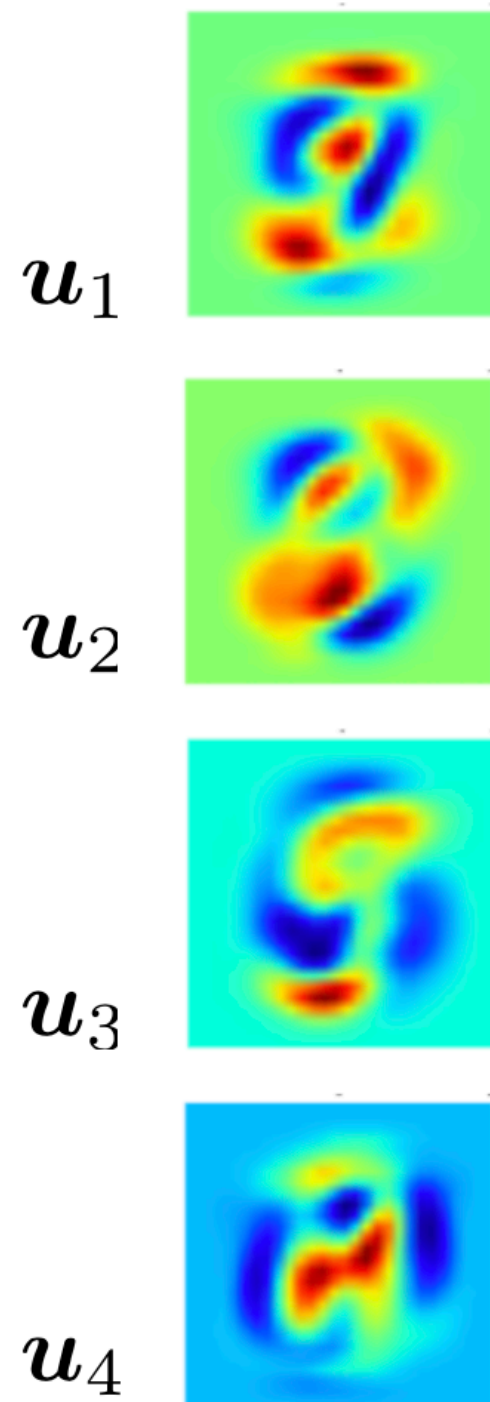
PCA via minimal reconstruction error

- ▶ We got: $\frac{1}{N} \sum_{n=1}^N \|\mathbf{x}_n - \tilde{\mathbf{x}}_n\|^2 = \sum_{i=M+1}^D \mathbf{u}_i^T \mathbf{S} \mathbf{u}_i = \sum_{i=M+1}^D \lambda_i$
- ▶ Solve for \mathbf{u}_i under constraint $\mathbf{u}_i^T \mathbf{u}_i = 1$ (else we get $\mathbf{u}_i = \mathbf{0}$)
- ▶ Method of Lagrange multipliers \rightarrow Solve eigensystem
 $\implies \mathbf{S} \mathbf{u}_i = \lambda_i \mathbf{u}_i$
- ▶ Find the largest M eigenvectors and values, such that the remaining $(D - M)$ are the smallest

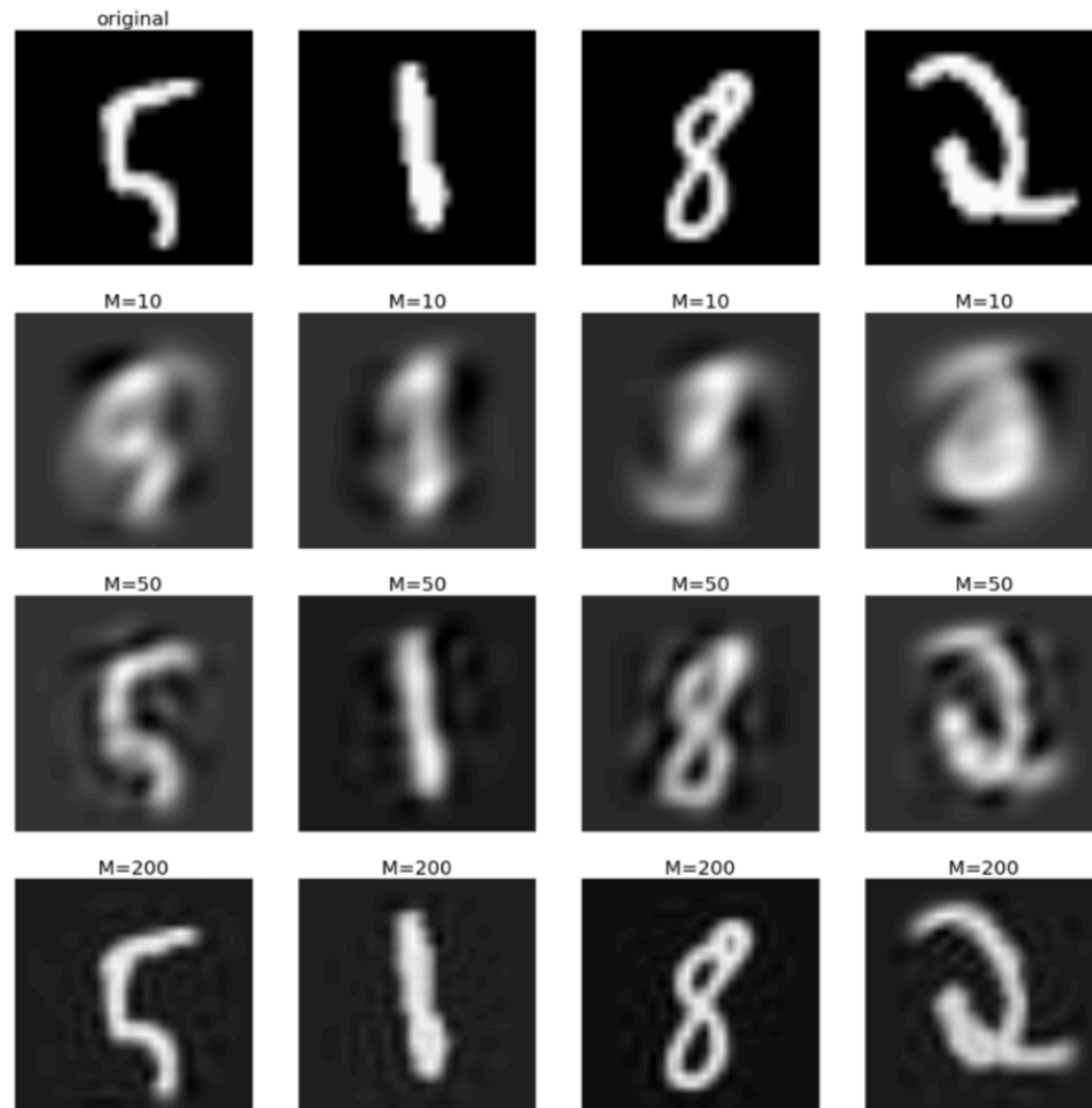
min reconstruction error = max projected variance

Applications: compression (MNIST)

Eigenvectors:



Reconstructions: $\tilde{x} = U_M z + \bar{x}$



Applications: compression (eigenfaces/UTKFace)

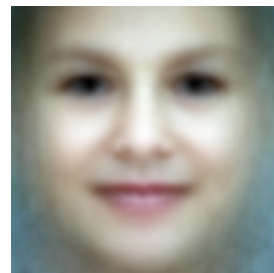
Dataset:



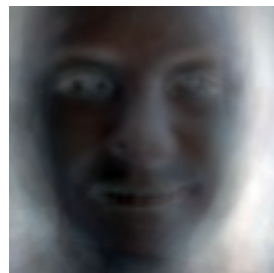
Task: Compress image

Method: Expand along principle components (PCA)

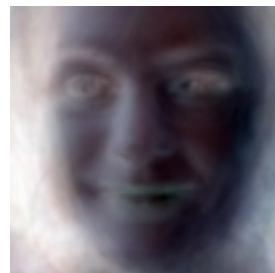
Mean $\bar{\mathbf{x}}$



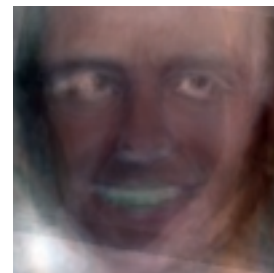
\mathbf{u}_1



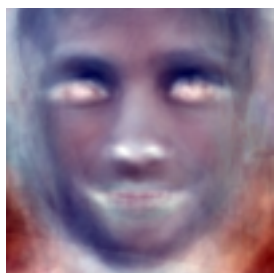
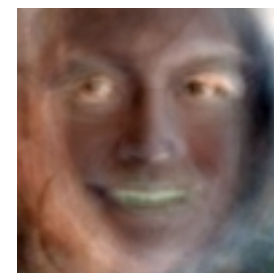
\mathbf{u}_2



\mathbf{u}_3



...



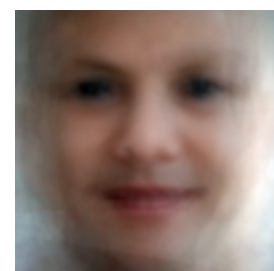
Result:

Original

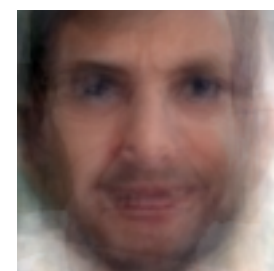


$$\approx \bar{\mathbf{x}} + \sum_{i=1}^M z_{ni} \mathbf{u}_i$$

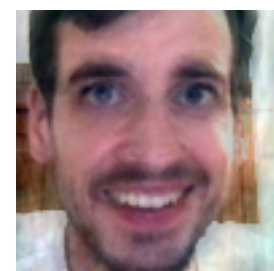
$M=1$



$M=10$



$M=50$



$M=150$

