CSE 575
Statistical Machine Learning

Lecture 19

YooJung Choi

Fall 2022

Principal component analysis

- Maximum variance formulation:

Maximize $\mathbf{u}_1^T \mathbf{\Sigma} \mathbf{u}_1$ s.t. $\|\mathbf{u}_1\|^2 = 1$, where $\mathbf{\Sigma} = \frac{1}{\mathrm{N}} \sum_{n=1}^{N} (\mathbf{x}_n - \boldsymbol{\mu})(\mathbf{x}_n - \boldsymbol{\mu})^T$

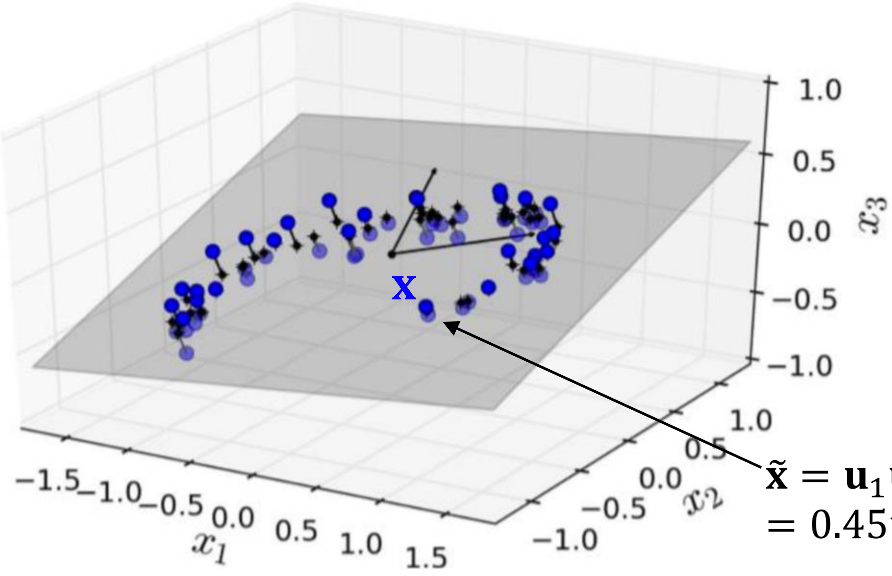
Maximize $\mathbf{u}_i^T \mathbf{\Sigma} \mathbf{u}_\mathrm{i}$ s.t. $\|\mathbf{u}_i\|^2 = 1$ and $\mathbf{u}_i^T \mathbf{u}_\mathrm{j} \;\; \forall j < i$, sequentially for each subsequent i

Principal component analysis

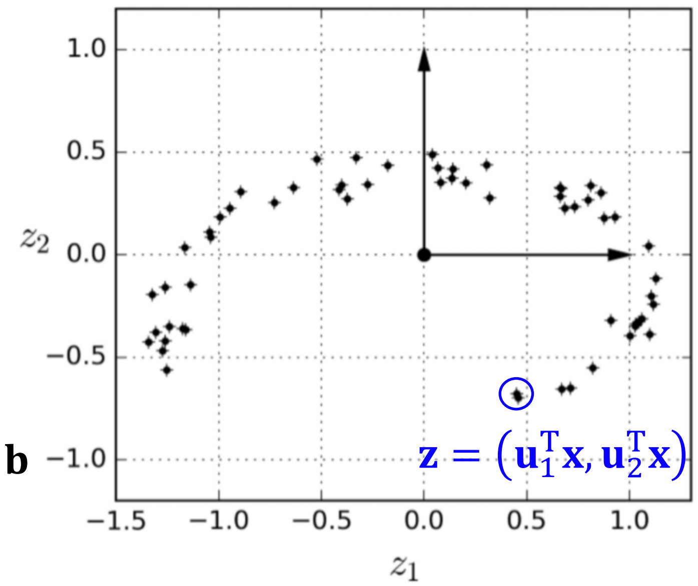
- Minimum error formulation: minimize the reconstruction error

$$J(\mathbf{x}, \tilde{\mathbf{x}}) = \frac{1}{N}\sum_{n=1}^{N} \|\mathbf{x}_n - \tilde{\mathbf{x}}_n\|^2$$

where $\tilde{\mathbf{x}}_n$ is a reconstruction from an M-dimensional latent space



$$\tilde{\mathbf{x}} = \mathbf{u}_1\mathbf{u}_1^{\mathrm{T}}\mathbf{x} + \mathbf{u}_2\mathbf{u}_2^{\mathrm{T}}\mathbf{x} + \mathbf{b}$$
$$= 0.45\mathbf{u}_1 - 0.7\mathbf{u}_2 + \mathbf{b}$$



$$\mathbf{z} = \left(\mathbf{u}_1^{\mathrm{T}}\mathbf{x}, \mathbf{u}_2^{\mathrm{T}}\mathbf{x}\right) = (0.45, -0.7)$$

Principal component analysis

- Minimum error formulation: minimize the reconstruction error

$$J(\mathbf{x}, \tilde{\mathbf{x}}) = \frac{1}{N} \sum_{n=1}^{N} \|\mathbf{x}_n - \tilde{\mathbf{x}}_n\|^2$$

  where $\tilde{\mathbf{x}}_n$ is a reconstruction from an M-dimensional latent space

- Let $\mathbf{w}_1, \dots, \mathbf{w}_D$ be orthonormal basis vectors. Then

$$\mathbf{x}_n = \sum_{i=1}^{D} (\mathbf{x}_n^T \mathbf{w}_i) \mathbf{w}_i$$

To show equivalence to maximum variance formulation:

1) Show $b_i = \boldsymbol{\mu}^T \mathbf{w}_i$ to minimize $J(\mathbf{x}, \tilde{\mathbf{x}})$
2) Show $J(\mathbf{x}, \tilde{\mathbf{x}}) = \sum_{i=M+1}^{D} \mathbf{w}_i^T \boldsymbol{\Sigma} \mathbf{w}_i$

- Reconstructed from an M-dimensional subspace,

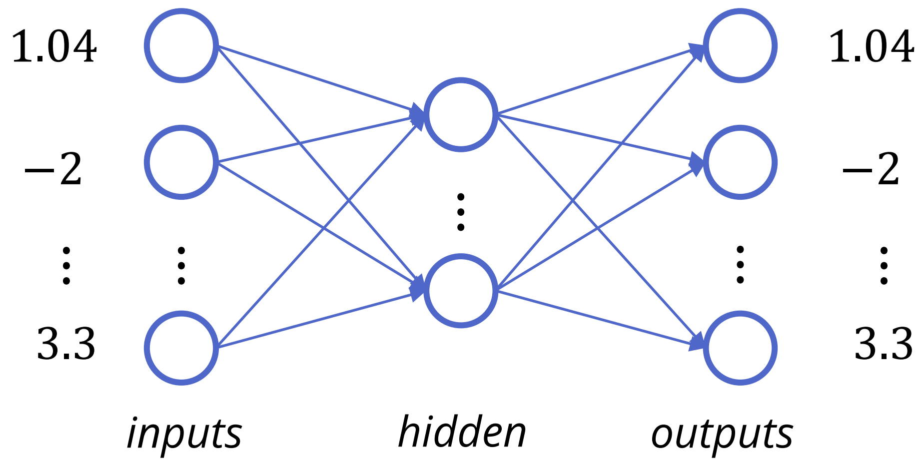
$$\tilde{\mathbf{x}}_n = \sum_{i=1}^{M} (\mathbf{x}_n^T \mathbf{w}_i) \mathbf{w}_i + \sum_{i=M+1}^{D} b_i \mathbf{w}_i$$

Linear in $\mathbf{x}_n$

Dim. Reduction with MLPs

- Consider an MLP with D input units, M hidden units, and D output units (M < D)
- Given an input $\mathbf{x}$, want the network to output $\mathbf{x}$ as closely as possible
- Minimize the sum-of-squares error:

$$E(\mathbf{W}) = \frac{1}{2}\sum_{n=1}^{N} \|\mathbf{y}(\mathbf{x}_n, \mathbf{W}) - \mathbf{x}_n\|^2$$



Dim. Reduction with MLPs

- Suppose the activation functions are linear
- Then $E(\mathbf{W}) = \frac{1}{2}\sum_{n=1}^{N}\|\mathbf{y}(\mathbf{x}_n, \mathbf{W}) - \mathbf{x}_n\|^2$ has a global minimum, which performs a projection onto the subspace spanned by the *M principal components*

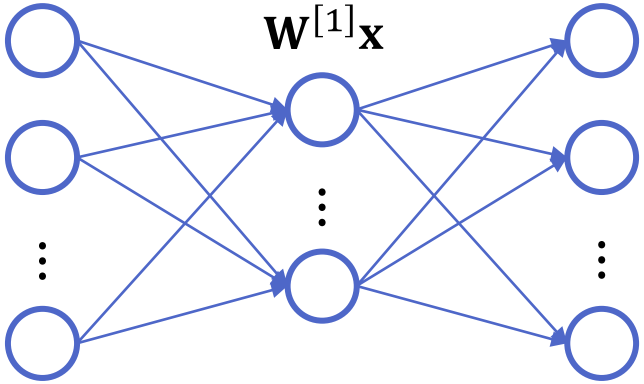
(True even with nonlinear hidden units)

M-dim projection

Note: the vectors are not necessarily orthogonal or normalized

$\mathbf{W}^{[1]}\mathbf{x}$

D-dim $\mathbf{x}$

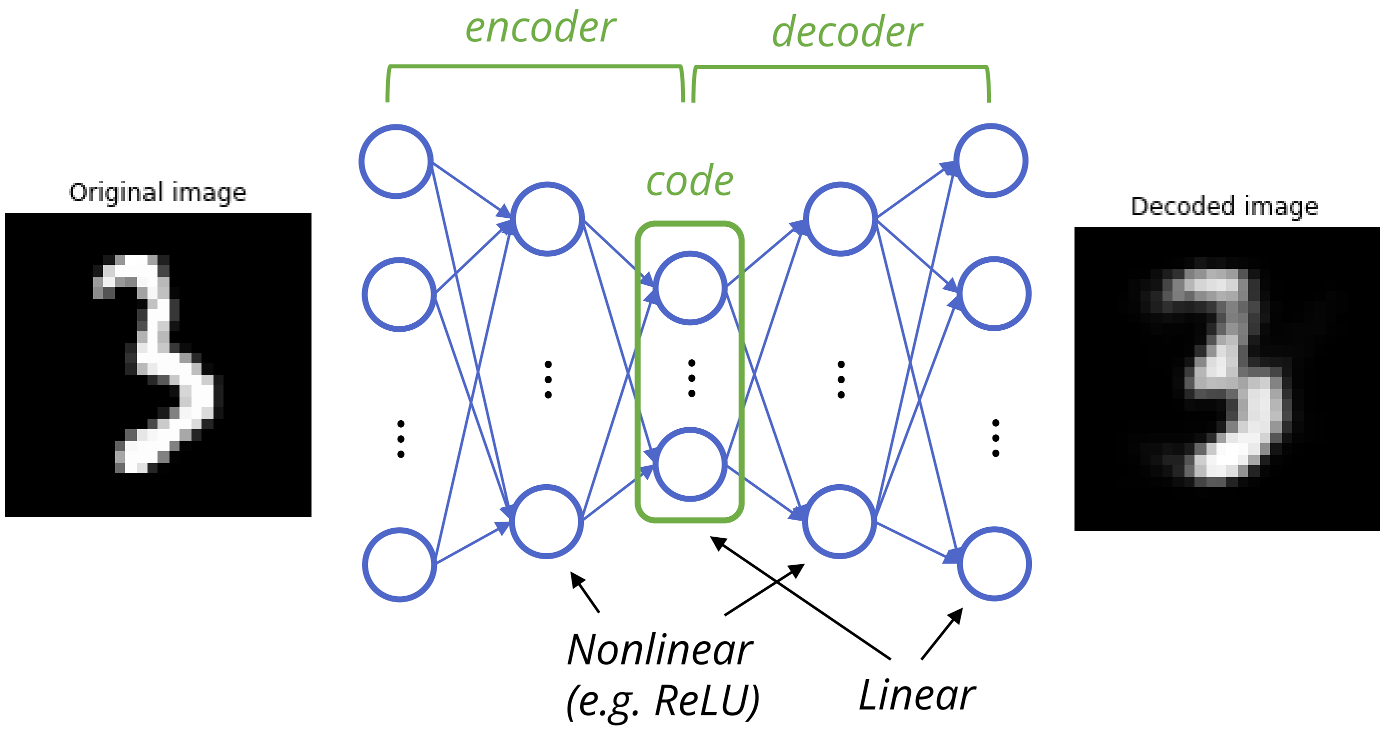


$\mathbf{y} = \mathbf{W}^{[2]}\left(\mathbf{W}^{[1]}\mathbf{x}\right)$ *D-dim reconstruction*

inputs *hidden* *outputs*

Autoencoders

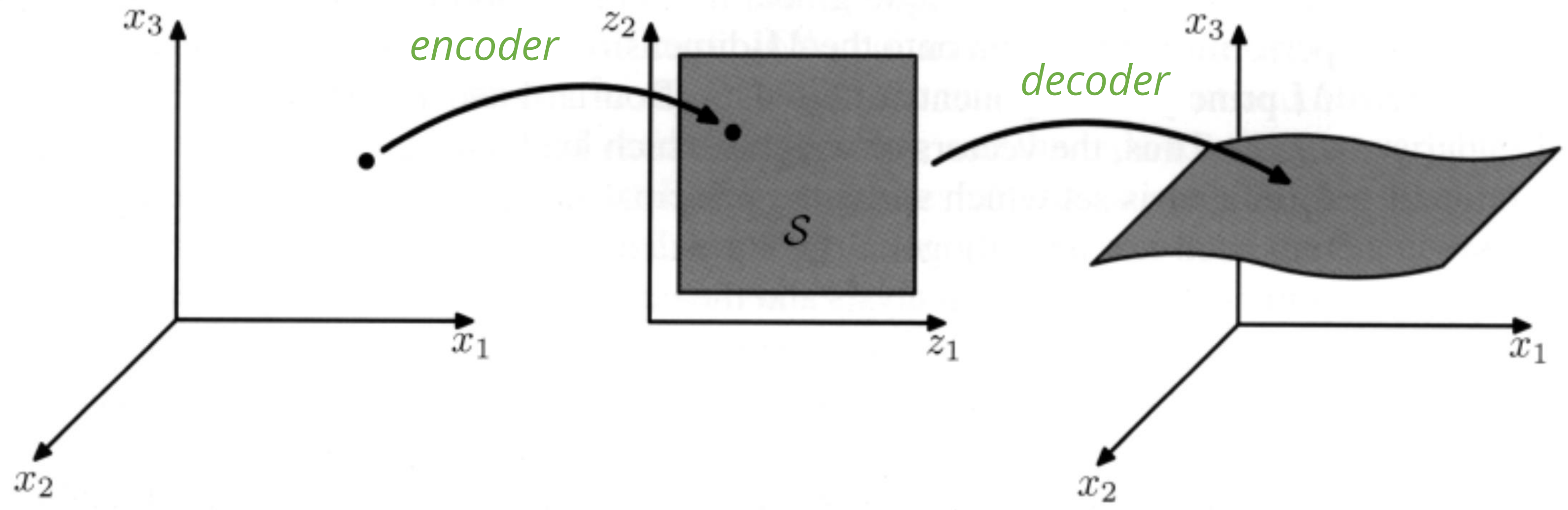
- With additional hidden layers (using nonlinear activation functions), the network performs *nonlinear dimensionality reduction*



Autoencoders

- With additional hidden layers (with nonlinear activation functions), the network performs *nonlinear dimensionality reduction*

$D = 3$
$M = 2$



Midterm 2 logistics

- Written exam, Wednesday 11/9, in-class
- Closed book.
- One single-sided letter-sized cheat sheet (with your name on it).
- No computers. Basic calculator allowed.
- Covers all materials. Focus on the following topics:
  - Support vector machines
  - Neural networks
  - Bayesian networks, Gaussian mixtures, EM
  - K-means
  - Dimension reduction

Project presentations

- Each group gets a 15-minute slot including Q&A (hard limit). Prepare to speak for 10-12 minutes. All members in your group need to present.
- Sign up for your presentation slot at https://links.asu.edu/CSE575-F22
- Submit your slides by noon on the day of presentation
- 10% of total grade. You will be graded on:
  - Problem description & motivation
  - Methodology used
  - Comparison with existing work
  - Lesson learned
  - Clarity of presentation
  - Q&A (both as presenters and audience)