CS540 Introduction to Artificial Intelligence
Lecture 12

Young Wu
Based on lecture slides by Jerry Zhu, Yingyu Liang, and Charles Dyer

July 13, 2020

Low Dimension Representation

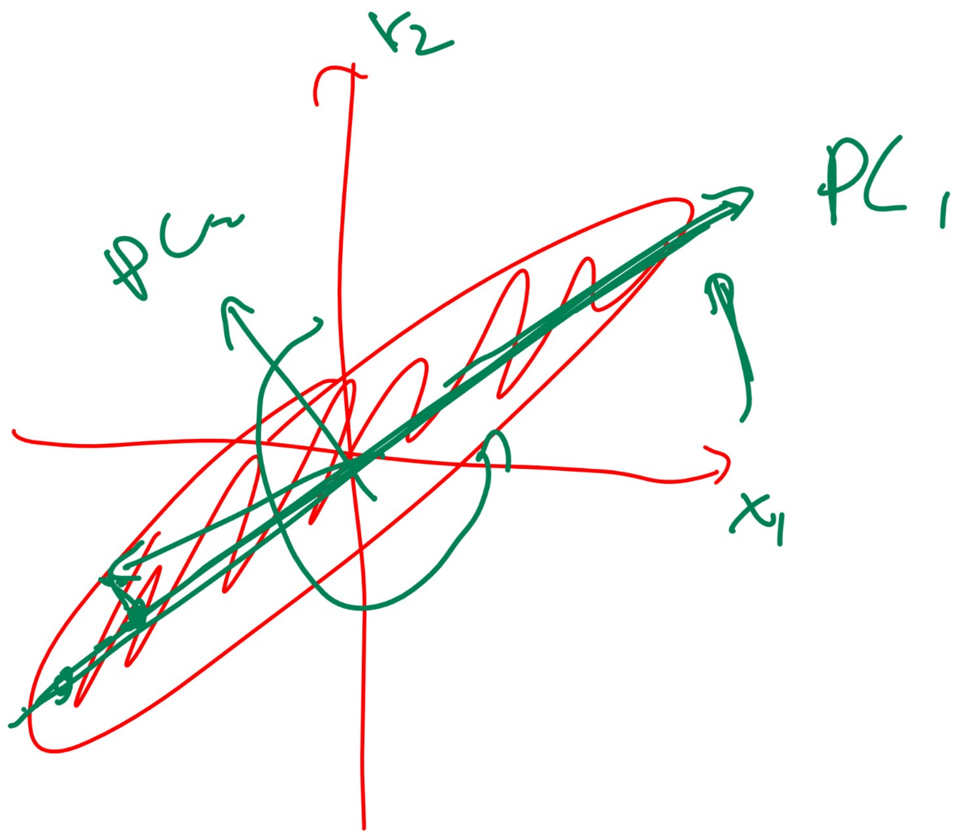
Motivation

- Unsupervised learning techniques are used to find low dimensional representation.

1. Visualization.
2. Efficient storage.
3. Better generalization.
4. Noise removal.

Dimension Reduction Diagram

Motivation



max $\underline{\text{variation}}$ along PC_1

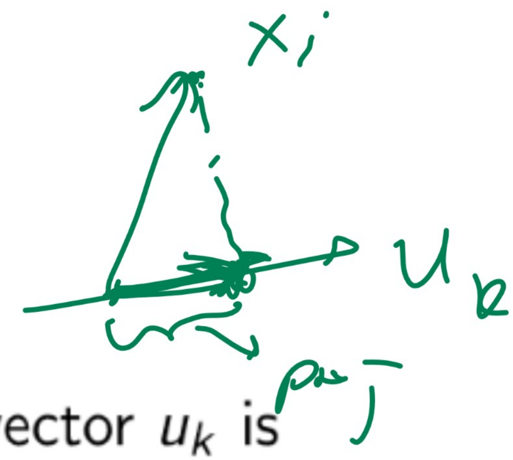
Projection

Definition

length

- The projection of x_i onto a unit vector u_k is the vector in the direction of u_k that is the closest to x_i.

$$\text{proj}_{u_k} x_i = \boxed{\left(\frac{u_k^T x_i}{u_k^T u_k}\right) u_k} = u_k^T x_i u_k$$



- The length of the projection of x_i onto a unit vector u_k is $u_k^T x_i$.

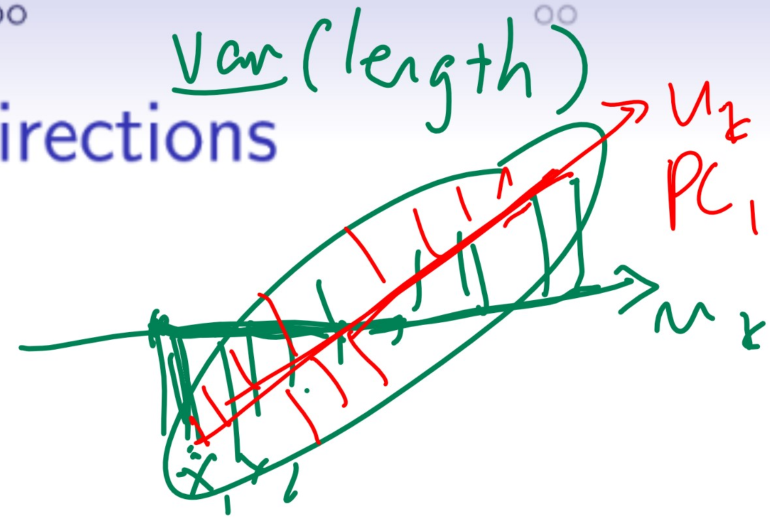
$$\left\| \text{proj}_{u_k} x_i \right\|_2 = \boxed{u_k^T x_i}$$

Maximum Variance Directions

Definition

$x^T A x \to A x$

$x^T x \to x$



- The goal is to find the direction that maximizes the projected variance.

variance

$$\max_{u_k} u_k^T \hat{\Sigma} u_k \text{ such that } u_k^T u_k = 1$$

Lagrange

$$\Rightarrow \max_{u_k} u_k^T \hat{\Sigma} u_k - \lambda u_k^T u_k$$

derivative.

$$\Rightarrow \hat{\Sigma} u_k = \lambda u_k$$

eigenvalue

max λ

$u_k^T \lambda u_k = \lambda$

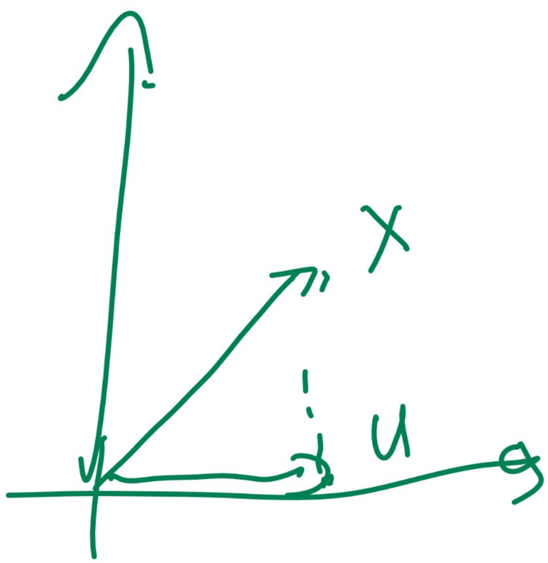
Projection Example 1

Quiz

- What is the projection of $\overset{x}{\begin{bmatrix}1\\1\end{bmatrix}}$ onto $\overset{u}{\begin{bmatrix}1\\0\end{bmatrix}}$?

$\text{proj} = \begin{pmatrix}1\\0\end{pmatrix}$

$$\frac{u^T x}{u^T u} u = \frac{1}{1}\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix}$$

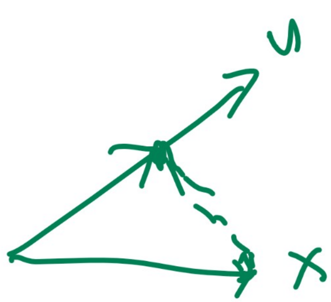


Projection Example 2

Quiz

- What is the projection of $\overset{x}{\begin{bmatrix}1\\0\end{bmatrix}}$ onto $\overset{u}{\begin{bmatrix}1\\1\end{bmatrix}}$?

$$\begin{pmatrix}\frac{1}{2}\\ \frac{1}{2}\end{pmatrix}$$



$$\frac{u^T x}{u^T u}\, u$$

$$= \frac{1}{2}\cdot\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}\frac{1}{2}\\ \frac{1}{2}\end{pmatrix}$$

Projection Example 3

Quiz

Q8

u

- What is the projection of $\begin{bmatrix}1\\2\\3\end{bmatrix}$ onto $\begin{bmatrix}1\\1\\1\end{bmatrix}$?

$\frac{u^T x u}{}$

not unit

- A: $[1 \quad 1 \quad 1]^T$
- B: $[2 \quad 2 \quad 2]^T$
- C: $[3 \quad 3 \quad 3]^T$
- D: $[4 \quad 4 \quad 4]^T$
- E: $[6 \quad 6 \quad 6]^T$

$$\frac{u^T x}{u^T u} u$$

$$= \frac{6}{3}\begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}2\\2\\2\end{pmatrix}$$

Projection Example 4

Quiz

Q9

- What is the projection of $\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$ onto $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$?
- A: $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$
- B: $\begin{bmatrix} 2 & 2 & 2 \end{bmatrix}^T$
- C: $\begin{bmatrix} 3 & 3 & 3 \end{bmatrix}^T$
- D: $\begin{bmatrix} 4 & 4 & 4 \end{bmatrix}^T$
- E: $\begin{bmatrix} 6 & 6 & 6 \end{bmatrix}^T$

Spectral Decomposition Example 1

Quiz

- Given the following spectral decomposition of $\hat{\Sigma}$, what is the first principal component?

eigen vectors — eigenvalues

$$\hat{\Sigma} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}^{-1}$$

$PC_1 \to$ largest e.v $\to 3 \Rightarrow$ normalized $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}$

$PC_2 \to \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ $\quad PC_3 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$

Spectral Decomposition Example 2

Quiz

Q10

- Given the following spectral decomposition of $\hat{\Sigma}$, what is the first principal component?

$$\hat{\Sigma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- A: $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, B: $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, C: $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, D: $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, E: $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

Spectral Decomposition Example 3

Quiz

Q11

- $\hat{\Sigma} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, what is the second principal component?
- A: $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, B: $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, C: $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, D: $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, E: $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

Principal Component Analysis

Algorithm

- Input: instances: $\{x_i\}_{i=1}^{n}$, the number of dimensions after reduction $K < m$.
- Output: K principal components.
- Find the largest K eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant ... \geqslant \lambda_K$.
- Return the corresponding unit orthogonal eigenvectors $u_1, u_2 ... u_K$.

Number of Dimensions

Discussion

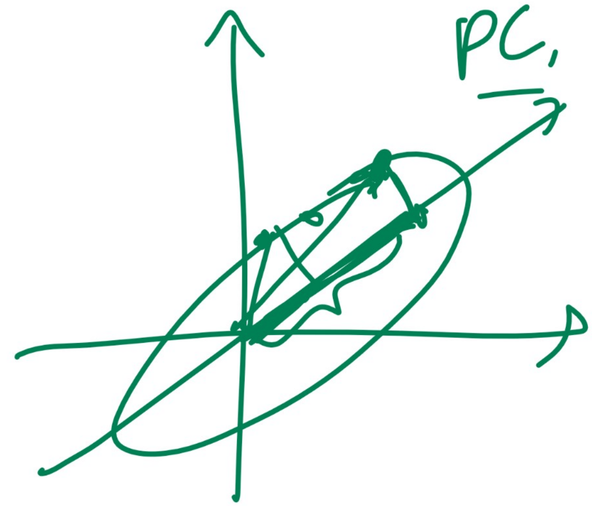
- There are a few ways to choose the number of principal components K.
- K can be selected given prior knowledge or requirement.
- K can be the number of non-zero eigenvalues.
- K can be the number of eigenvalues that are large (larger than some threshold).

> 1

PCA

Reduced Feature Space

Discussion



- The original feature space is m dimensional.

$$(x_{i1}, x_{i2}, \dots, x_{im})^T$$

- The new feature space is K dimensional.

$$\left(u_1^T x_i, u_2^T x_i, \dots, u_K^T x_i\right)^T$$

- Other supervised learning algorithms can be applied on the new features.

Eigenface

Discussion

in m dim

- Eigenfaces are eigenvectors of face images (pixel intensities or HOG features).
- Every face can be written as a linear combination of eigenfaces. The coefficients determine specific faces.

$$\boxed{x_i = \sum_{k=1}^{m} \left(u_k^T x_i\right) u_k} \approx \sum_{k=1}^{K} \left(u_k^T x_i\right) u_k$$

new features

- Eigenfaces and SVM can be combined to detect or recognize faces.

Reduced Space Example 1

Quiz

$u_3 \cdot x = 3$

$u_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

- 2017 Fall Final Q10
- If $u_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$ and $u_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$. If one original data is $x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. What is the new representation?

$$\begin{pmatrix} u_1 \cdot x \\ u_2 \cdot x \end{pmatrix} = \begin{pmatrix} \frac{3}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

Reduced Space Example 1 Diagram

Quiz

$$X \approx \frac{3}{\sqrt{2}}\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\\ 0\end{pmatrix} + \frac{1}{\sqrt{2}}\begin{pmatrix}-\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\\ 0\end{pmatrix} + 3\begin{pmatrix}0\\ 0\\ 1\end{pmatrix}$$

$$= \begin{pmatrix}\frac{3}{2}\\ \frac{3}{2}\\ 0\end{pmatrix} + \begin{pmatrix}-\frac{1}{2}\\ \frac{1}{2}\\ 0\end{pmatrix} = \begin{pmatrix}1\\ 2\\ 0\end{pmatrix} \approx \begin{pmatrix}1\\ 2\\ 3\end{pmatrix}$$

$$+\begin{pmatrix}0\\ 0\\ 3\end{pmatrix} = \begin{pmatrix}1\\ 2\\ 3\end{pmatrix} = \underline{X}$$

Reduced Space Example 2

Quiz

Q12

$PC_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, $PC_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$

- $\hat{\Sigma} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. If one original data is $x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. What is the new representation using only the first two principal components?
- A: $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, B: $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$, C: $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$, D: $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$, E: $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$

Reduced Space Example 3

Quiz

Q13, (last)

- $\hat{\Sigma} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. If one original data is $x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. What is the reconstructed vector using only the first two principal components?
- A: $\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$, B: $\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$, C: $\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$, D: $\begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$, E: $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

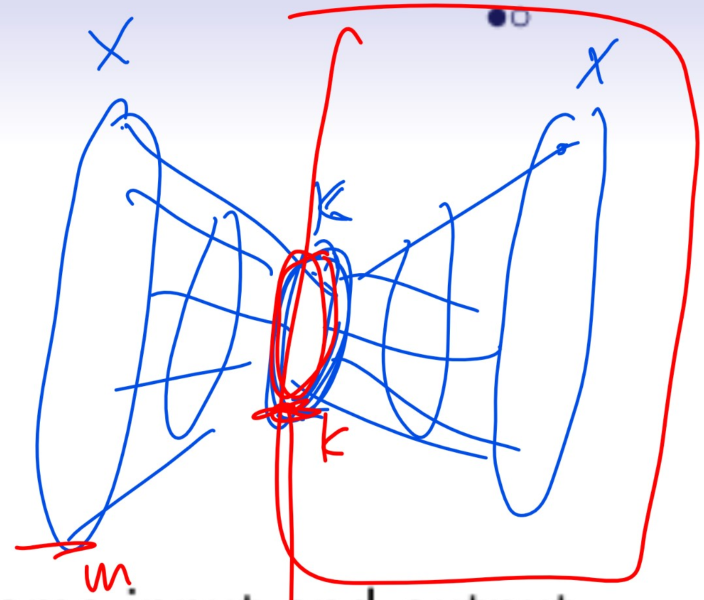
$$2\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 1\cdot\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$$

Autoencoder

Discussion

PCA $\simeq$ no activation $a = x^T w + b$

$a = g(x^T w + b)$



non-linear PCA

- A multi-layer neural network with the same input and output $y_i = x_i$ is called an autoencoder.
- The hidden layers have fewer units than the dimension of the input m.
- The hidden units form an encoding of the input with reduced dimensionality.

Kernel PCA

Discussion

- A kernel can be applied before finding the principal components.

$$\hat{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n} \underbrace{\varphi(x_i)}_{\infty \text{ dim}} \varphi(x_i)^T$$

- The principal components can be found without explicitly computing $\varphi(x_i)$, similar to the kernel trick for support vector machines.
- Kernel PCA is a non-linear dimensionality reduction method.