

선형대수 Term Project

SVD and PCA

12/14: Video record lecture

12/16: Live zoom lecture for Q/A



Introduction

1. Implement SVD for randomly generated 100-D vectors
2. Find some principal component
3. Represent 100-D vectors with the selected principal component vectors
- 4 Discuss the errors in representing vectors with partial basis (less than 20~30 basis vectors)

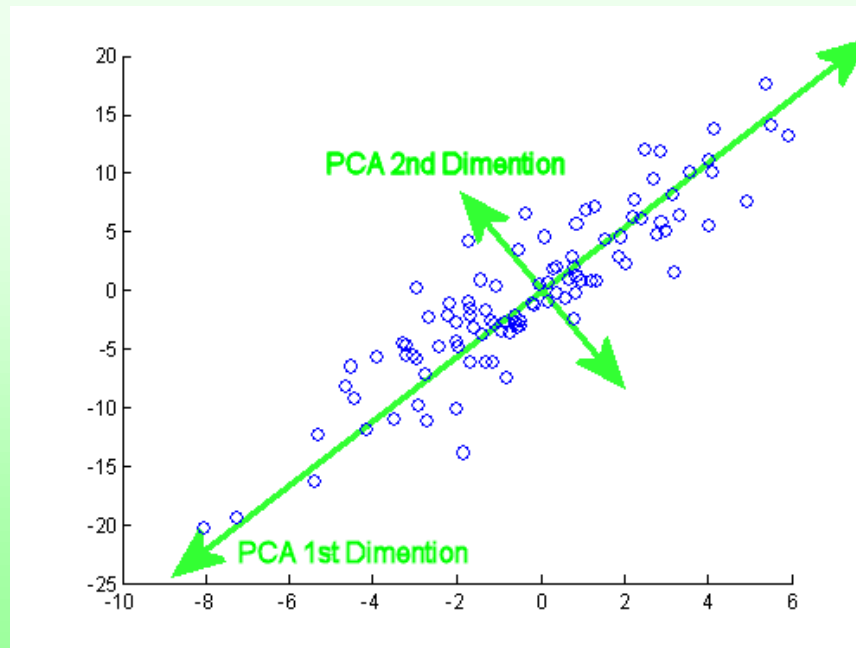


Principal Component Analysis

□ Correlation/Covariance matrix : $R = A^T A$

- The origin should be included in the vector space

□ The larger is singular value, the more dominant is the basis vector.



Generate Matrix A (1/2)

□ Generate 100-D vectors with random number

- $[x_1, x_2, x_3, \dots, x_{100}]$
- Define random number range for elements
- 1000 vectors are generated

□ Design specific relations between 2 selected elements

- Ex: $x_1 = x_2, x_5 = -2x_7$
- Random number range: $-100 \sim +100$
- Define 5 relations between 2 elements
 - Total 10 elements are selected

□ The other elements

- 45 elements: Random number range: $-20 \sim +20$
- 45 elements: $-5 \sim +5$



Generate Matrix A (2/2)

- Goal: principal components generation
 - Some elements are widely distributed
- Example for 5-D vectors generation
 - $x_1 = 2x_2$: x_2 range $-100 \sim 100$
 - x_3, x_4 : range $-20 \sim 20$
 - x_5 : range $-5 \sim 5$
- Generated vectors
 - $[90, 45, -12, 14, 2]$
 - $[-120, -60, 16, -9, -1]$
 - $[28, 14, -2, 18, 0]$



SVD and PCA

□ Generated 100-D vectors in row of matrix A

- A : 1000x100 matrix
- $A^T A$: 100x100 matrix
- Basis of 100-D vectors are in matrix U

□ Select principal components

- Select the principal components (eigenvectors) with the largest singular values



Vector Representation

□ Represent 100-D vectors

- Increase the number of principal components from 10, 15, 20, 25, \dots , rank (A)
- 100 randomly generated 100-D vectors

□ Check the errors in representation

- Vector distances between the generated vector and represented vectors with the basis vector.
- Plot the graph for the average distance of 100 vectors w.r.t the number of basis vectors
 - X-axis: 10, 15, 20, 25, 30 , \dots .rank(A)
 - Y-axis: average distance for each basis



Report Submission

□ Due: 12/19 (Sunday) 21:00

- Upload pdf file to LMS homework

□ Report

- Ppt type 15~20 pages
- Describe your definition and discussion
- No Presentation
- No codes explanation in the report

□ Use library

- C/C++/Python(numpy)/matlab ...
- You can use fully the functions and classes in the open sources.

