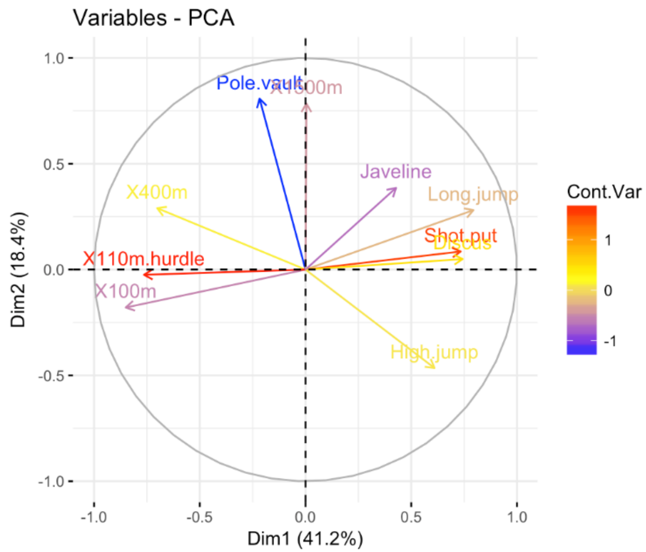
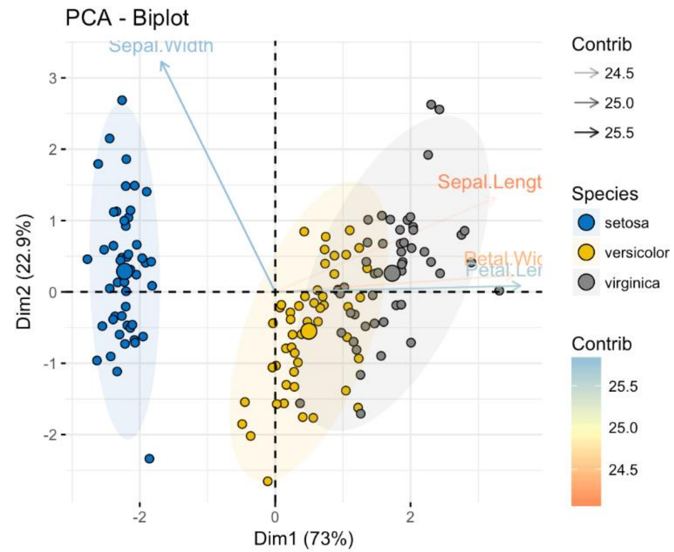


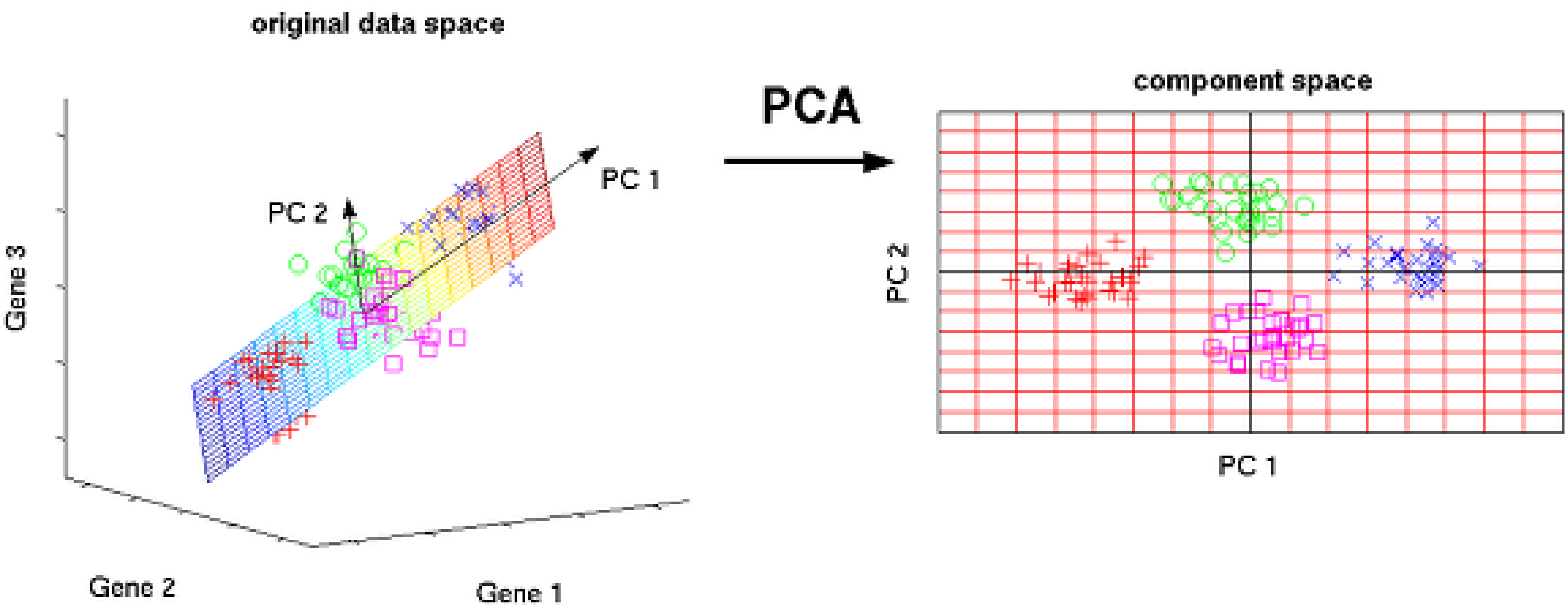


Dimensionality Reduction: Principal Component Analysis

Pilsung Kang
School of Industrial Management Engineering
Korea University

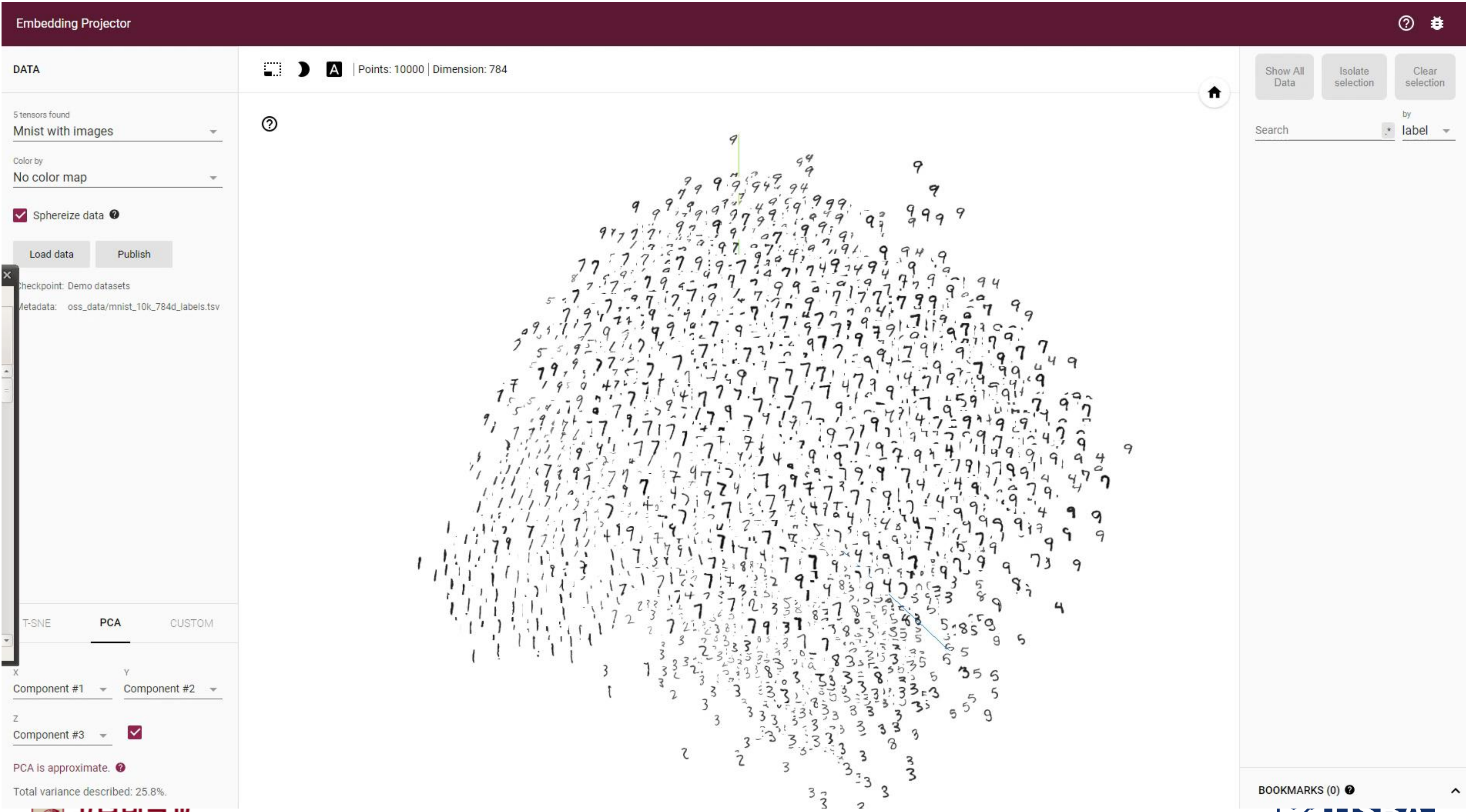
Principal Component Analysis: PCA

- Principal Component Analysis: PCA
  - ✓ To find a set orthogonal bases to preserve the variance of the original data



Principal Component Analysis: PCA

- PCA Example: MNIST Dataset

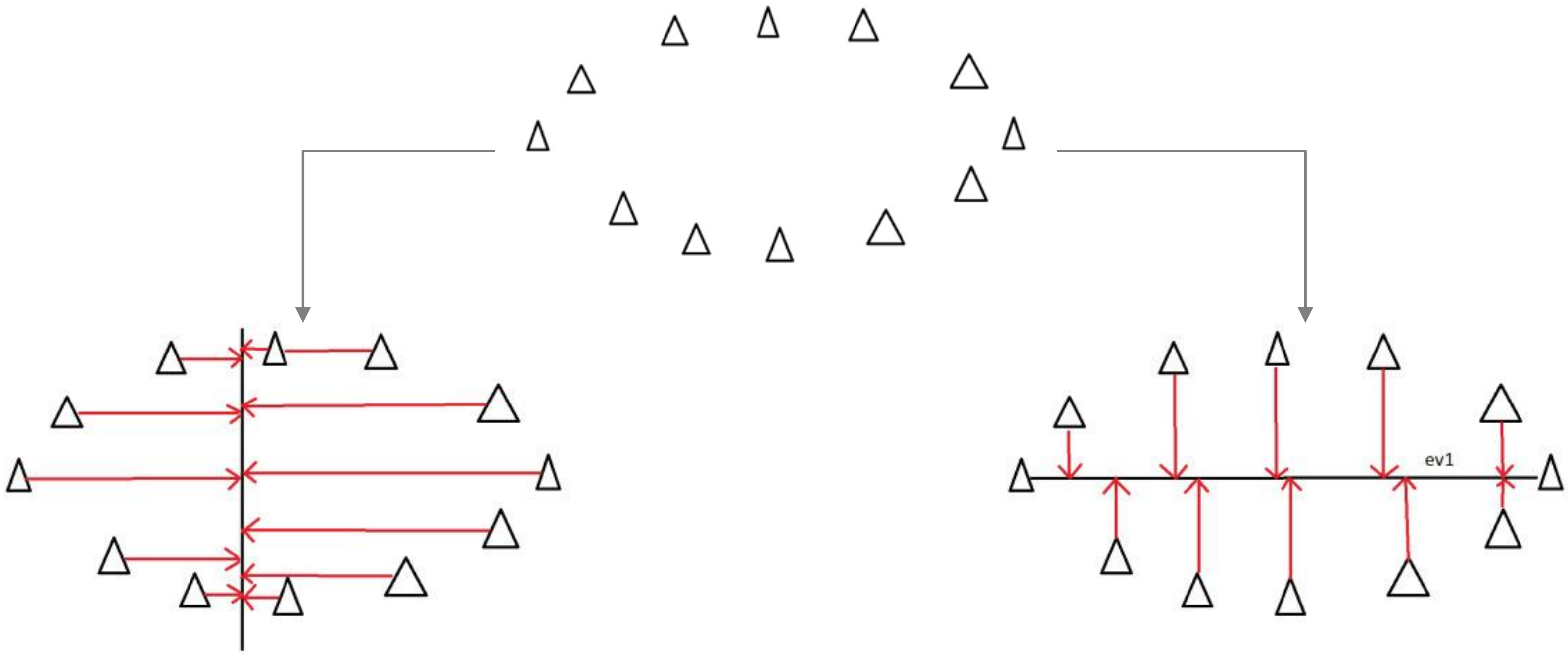


Principal Component Analysis: PCA

- Paintings

Principal Component Analysis: PCA

- Which bases are preferred?



Principal Component Analysis: PCA

- Purpose
  - ✓ Find a set of basis that can preserve the variance as much as possible after the projection on the basis
    - $X_1, X_2, \ldots, X_p$: Original variables
    - $\mathbf{a}_i = [a_{i1}, a_{i2}, \ldots, a_{ip}]$: i^{th} basis or principal component
    - $Y_1, Y_2, \ldots, Y_p$:Variables after the projection onto the i^th basis

$$Y_1 = \mathbf{a}_1^{'}\mathbf{X} = a_{11}X_1 + a_{12}X_2 + \cdots a_{1p}X_p$$

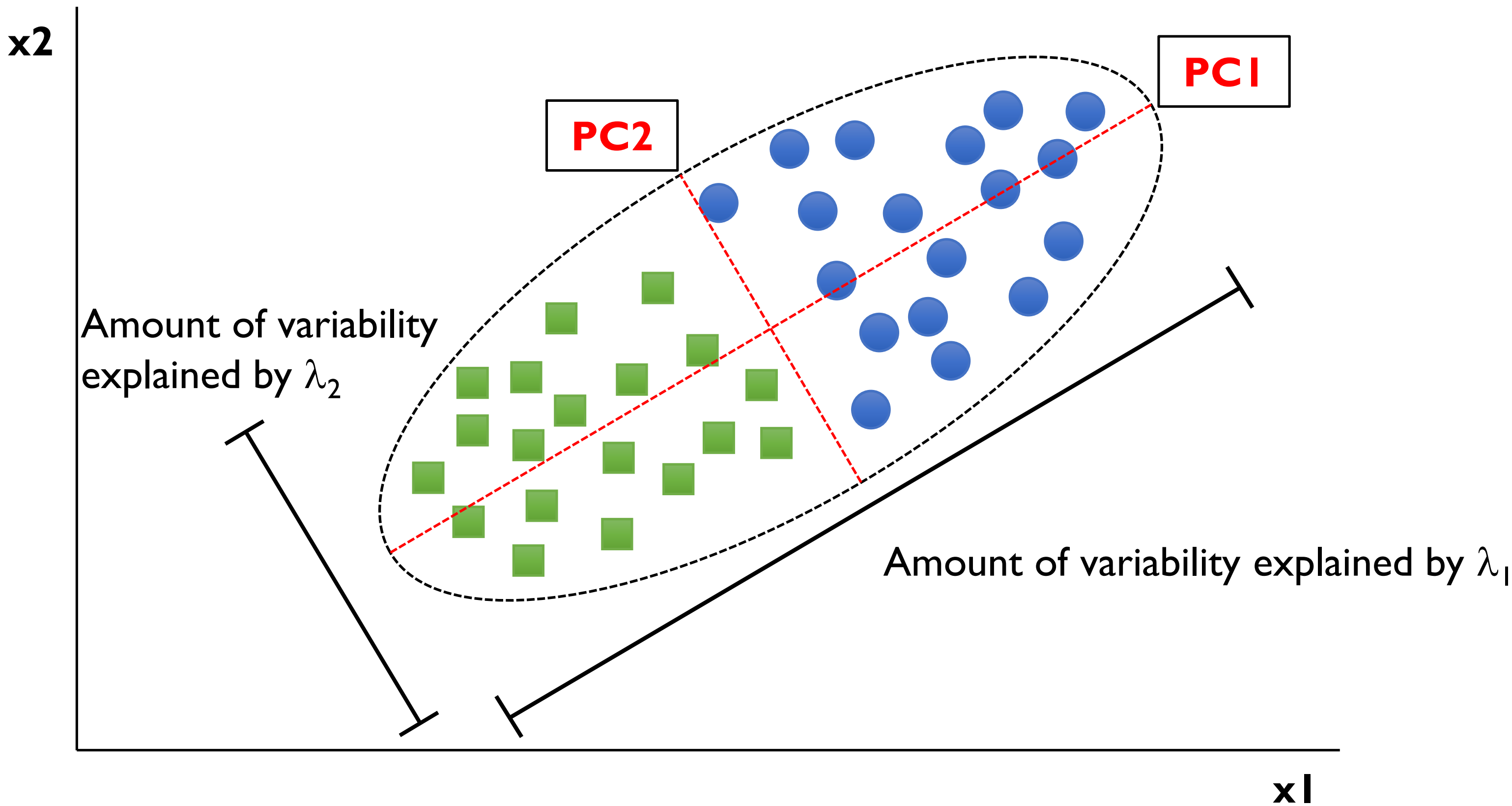
$$Y_2 = \mathbf{a}_2^{'}\mathbf{X} = a_{21}X_1 + a_{22}X_2 + \cdots a_{2p}X_p$$

$$\vdots \qquad\qquad\qquad \vdots$$

$$Y_p = \mathbf{a}_p^{'}\mathbf{X} = a_{p1}X_1 + a_{p2}X_2 + \cdots a_{pp}X_p$$

Principal Component Analysis: PCA

- How much variance can be preserved?

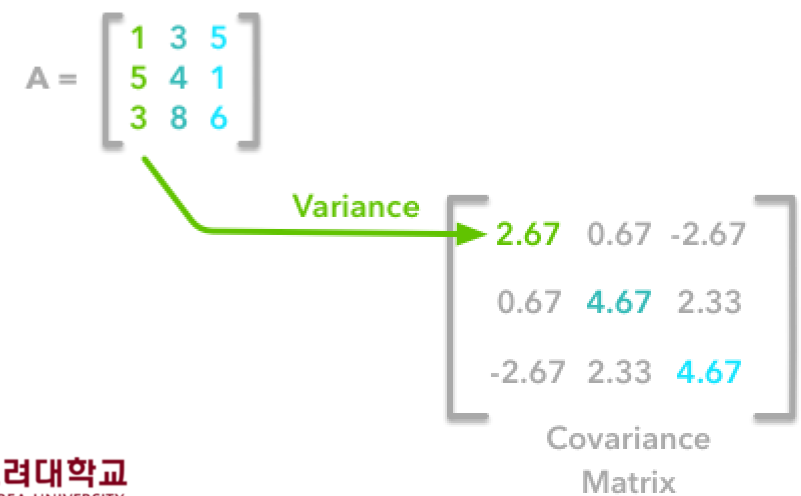


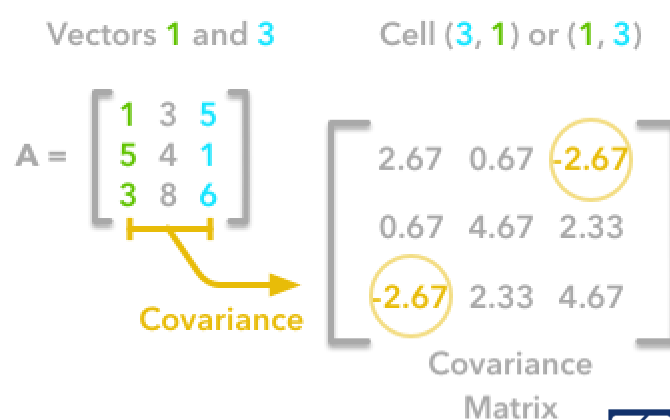
Principal Component Analysis: PCA

- Principal Component Analysis: PCA
  - ✓ Covariance
    - $\mathbf{X}$: a data set (d by n, d: # of variables, n: # of records)

$$Cov(\mathbf{X}) = \frac{1}{n}(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^T$$

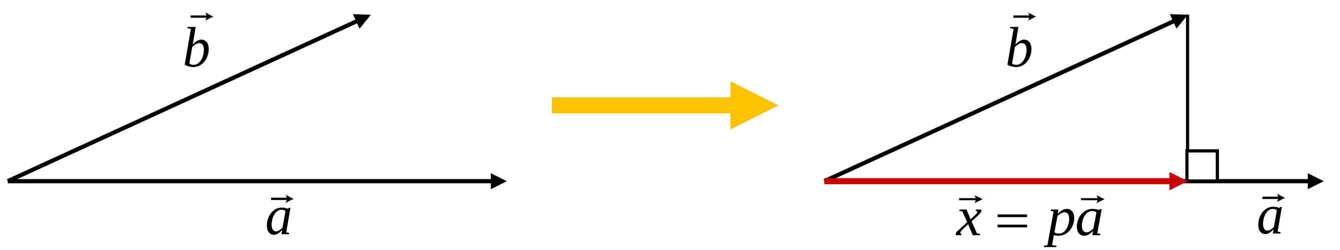
    - $\mathrm{Cov}(\mathbf{X})_{ij} = \mathrm{Cov}(\mathbf{X})_{ji}$
    - Total variance of the data set: $\mathrm{tr}[\mathrm{Cov}(\mathbf{X})] = \mathrm{Cov}(\mathbf{X})_{11} + \mathrm{Cov}(\mathbf{X})_{22} + \mathrm{Cov}(\mathbf{X})_{33} + \ldots + \mathrm{Cov}(\mathbf{X})_{dd}$





Principal Component Analysis: PCA

- Principal Component Analysis: PCA
  - ✓ Projection onto a basis



$$(\vec{b} - p\vec{a})^{\mathrm{T}}\vec{a} = 0 \quad \Rightarrow \quad \vec{b}^{\mathrm{T}}\vec{a} - p\vec{a}^{\mathrm{T}}\vec{a} = 0 \quad \Rightarrow \quad p = \frac{\vec{b}^{\mathrm{T}}\vec{a}}{\vec{a}^{\mathrm{T}}\vec{a}}$$

$$\vec{x} = p\vec{a} = \frac{\vec{b}^{\mathrm{T}}\vec{a}}{\vec{a}^{\mathrm{T}}\vec{a}}\vec{a}$$

If $\vec{a}$ is unit vector

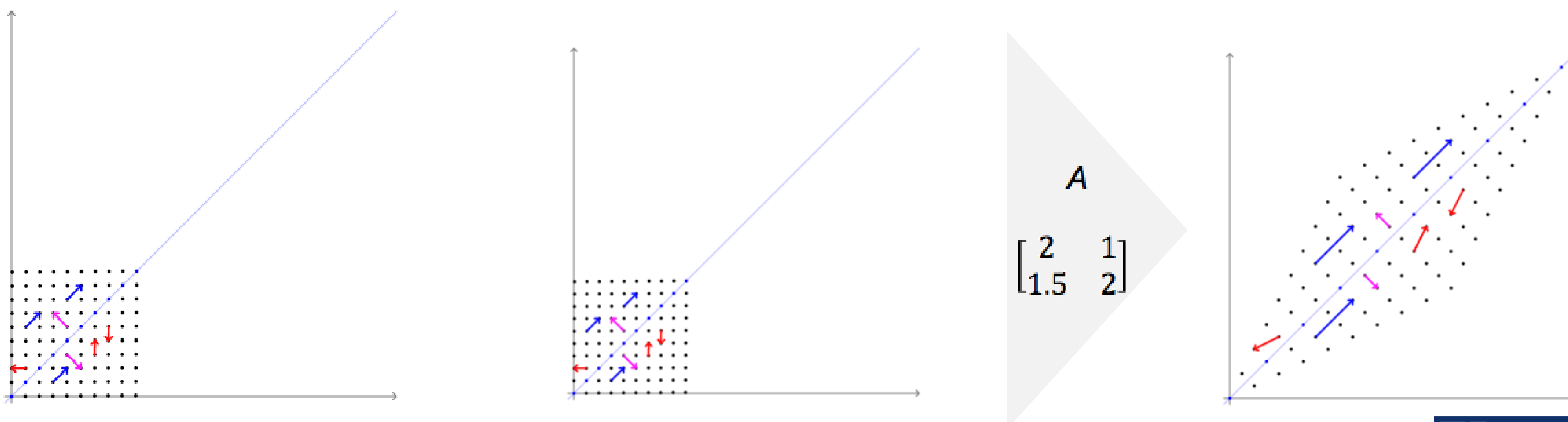
$$p = \vec{b}^{\mathrm{T}}\vec{a} \quad \Rightarrow \quad \vec{x} = p\vec{a} = (\vec{b}^{\mathrm{T}}\vec{a})\vec{a}$$

Principal Component Analysis: PCA

- Principal Component Analysis: PCA
  - ✓ Eigenvalue and eigenvector
    - When matrix A is given, scalar value λ and vector $\mathbf{x}$ that satisfy the following equation are called eigenvalue and eigenvector, respectively.

$$\mathbf{Ax} = \lambda\mathbf{x} \quad \rightarrow \quad (\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = 0$$

  - ✓ Multiplying a matrix to a vector means a linear transformation is conducted.
    - Eigenvectors do not change the direction by the transformation



Principal Component Analysis: PCA

- Principal Component Analysis: PCA
  - ✓ Eigenvalue and eigenvector
  - ✓ If a matrix A is non-singular d by d matrix,
    - There exist d eigenvalue-eigenvector pairs
    - Eigenvectors are orthogonal to each other

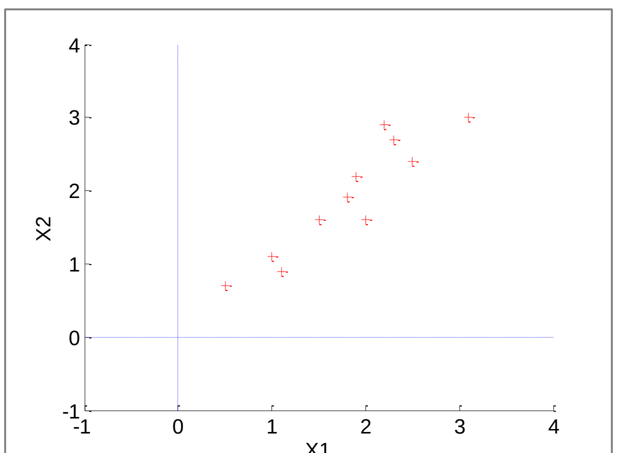
$$\mathrm{tr}(\mathbf{A}) = \lambda_1 + \lambda_2 + \cdots \lambda_d$$

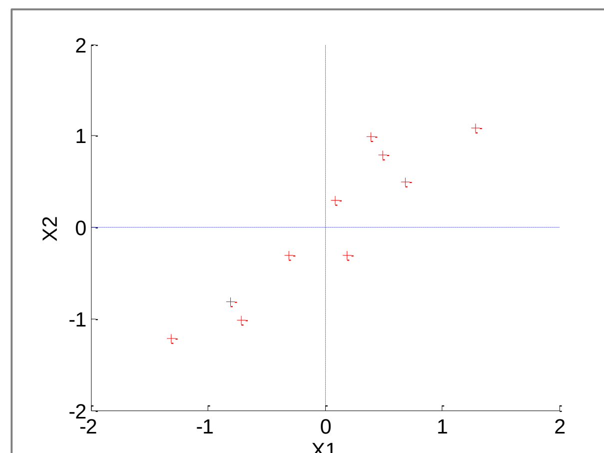
PCA Procedure

- Step 1: Data Centering
  - ✓ Make the mean of the variables equal to 0

x_1	2.5	0.5	2.2	1.9	3.1	2.3	2	1	1.5	1.1
x_2	2.4	0.7	2.9	2.2	3	2.7	1.6	1.1	1.6	0.9

x_1	0.69	-1.31	0.39	0.09	1.29	0.49	0.19	-0.81	-0.31	-0.71
x_2	0.49	-1.21	0.99	0.29	1.09	0.79	-0.31	-0.81	-0.31	-1.01





PCA Procedure

- Step 2: Formulate the optimization problem
  - ✓ If a vector x is projected onto a basis w, then the variance after the projection becomes

$$\mathrm{V} = \frac{1}{n}(\mathbf{w}^{\mathrm{T}}\mathbf{X})(\mathbf{w}^{\mathrm{T}}\mathbf{X})^{\mathrm{T}} = \frac{1}{n}\mathbf{w}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{w} = \mathbf{w}^{\mathrm{T}}\mathbf{S}\mathbf{w}$$

    - S is the sample covariance matrix where **x** is normalized.

  - ✓ The purpose of PCA is to maximize the variance V after projection

$$\max \mathbf{w}^{\mathrm{T}}\mathbf{S}\mathbf{w}$$

$$\mathrm{s.t.}\ \mathbf{w}^{\mathrm{T}}\mathbf{w} = 1$$

$$S = \begin{pmatrix} 0.6166 & 0.6154 \\ 0.6154 & 0.7166 \end{pmatrix}$$

고려대학교 KOREA UNIVERSITY

DSBA Data Science & Business Analytics

PCA Procedure

- Step 3: Obtain the solution
  - ✓ By employing Lagrangian multiplier,

$$\max \quad \mathbf{w}^T \mathbf{S} \mathbf{w}$$

$$s.t. \quad \mathbf{w}^T \mathbf{w} = 1$$

$$L = \mathbf{w}^T \mathbf{S} \mathbf{w} - \lambda(\mathbf{w}^T \mathbf{w} - 1)$$

$$\frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{S}\mathbf{w} - \lambda \mathbf{w} = 0 \ \Rightarrow (\mathbf{S} - \lambda \mathbf{I})\mathbf{w} = 0$$

$$Eigenvectors = \begin{bmatrix} 0.6779 & -0.7352 \\ 0.7352 & 0.6779 \end{bmatrix} \qquad Eigenvalues = (1.2840 \quad 0.0491)$$

고려대학교 KOREA UNIVERSITY

DSBA Data Science & Business Analytics

PCA Procedure

- Step 4: Find the base set of bases

  ✓ Sort the eigenvectors in a descending order of eigenvalues

$$Eigenvectors = \begin{bmatrix} 0.6779 & -0.7352 \\ 0.7352 & 0.6779 \end{bmatrix} \qquad Eigenvalues = (1.2840 \quad 0.0491)$$

  ✓ Let $\mathbf{w}_1$ be one of the eigenvectors and λ_1 be the corresponding eigenvalue.

  ✓ The variation of the samples projected onto $\mathbf{w}_1$ is

$$\mathbf{v} = (\mathbf{w}_1^T \mathbf{X})(\mathbf{w}_1^T \mathbf{X})^T = \mathbf{w}_1^T \mathbf{X}\mathbf{X}^T \mathbf{w}_1 = \mathbf{w}_1^T \mathbf{S}\mathbf{w}_1$$

$$\text{Since} \quad \mathbf{S}\mathbf{w}_1 = \lambda_1 \mathbf{w}_1, \quad \mathbf{w}_1^T \mathbf{S}\mathbf{w}_1 = \mathbf{w}_1^T \lambda_1 \mathbf{w}_1 = \lambda_1 \mathbf{w}_1^T \mathbf{w}_1 = \lambda_1$$

    - One basis can preserve 96% of the original variance in this example (1.2840/(0.0491+1.2840))

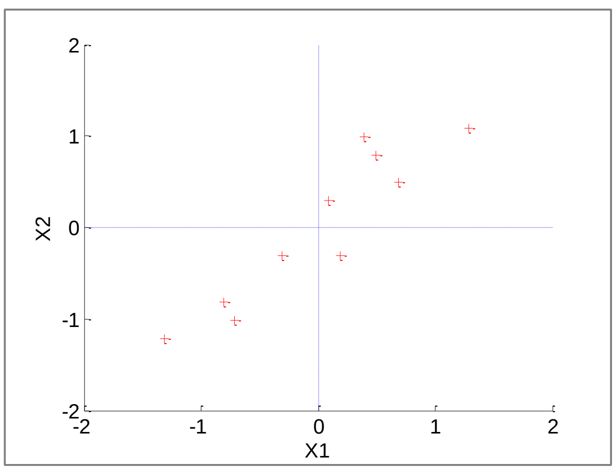
고려대학교 KOREA UNIVERSITY

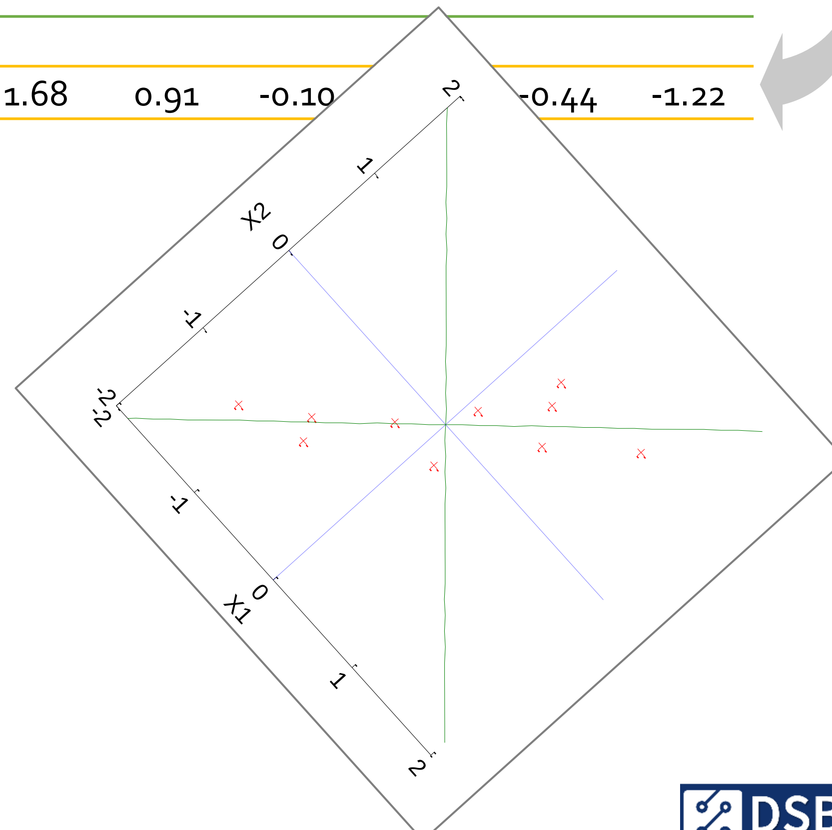
DSBA Data Science & Business Analytics

PCA Procedure

- Step 5: Extract new features
  - ✓ Project the original data onto the selected bases

x_1	0.69	-1.31	0.39	0.09	1.29	0.49	0.19	-0.81	-0.31	-0.71
x_2	0.49	-1.21	0.99	0.29	1.09	0.79	-0.31	-0.81	-0.31	-1.01
z_1	0.83	-1.78	0.99	0.27	1.68	0.91	-0.10	[illegible]	-0.44	-1.22





PCA Procedure

- Step 6: Reconstruct the original data
  - ✓ Reconstruct the data from the projected space into the original space

$$\mathbf{X} \xrightarrow{\text{Projection}} \mathbf{w}^T\mathbf{X} \xrightarrow{\text{Reconstruction}} \mathbf{w}\mathbf{w}^T\mathbf{X}$$

(d by n) (1 by d) (d by n) (d by 1)(1 by d) (d by n)

$\mathbf{x}_1$	0.69	-1.31	0.39	0.09	1.29	0.49	0.19	-0.81	-0.31	-0.71
$\mathbf{x}_2$	0.49	-1.21	0.99	0.29	1.09	0.79	-0.31	-0.81	-0.31	-1.01
$\mathbf{z}_1$	0.83	-1.78	0.99	0.27	1.68	0.91	-0.10	-1.14	-0.44	-1.22
$\mathbf{x}'_1$	0.56	-1.21	0.67	0.19	1.14	0.62	-0.07	-0.78	-0.30	-0.83
$\mathbf{x}'_2$	0.61	-1.31	0.73	0.20	1.23	0.67	-0.07	-0.84	-0.32	-0.90

PCA Issue

- How many principal components are optimal?
  - ✓ No explicit solution
  - ✓ Can be determined based on the variance preservation ratio and domain expert's knowledge
  - ✓ Based on the properties of covariance, eigenvalue-eigenvectors,
    - Total variable of the dataset = Sum of eigenvalues of the sample covariance matrix

$$\begin{aligned} Total\ population\ variance &= \sigma_{11} + \sigma_{22} + \ldots + \sigma_{dd} \\ &= \lambda_1 + \lambda_2 + \ldots + \lambda_d \end{aligned}$$

  - ✓ If we project the data onto the k^{th} basis, the amount of variance preserved is

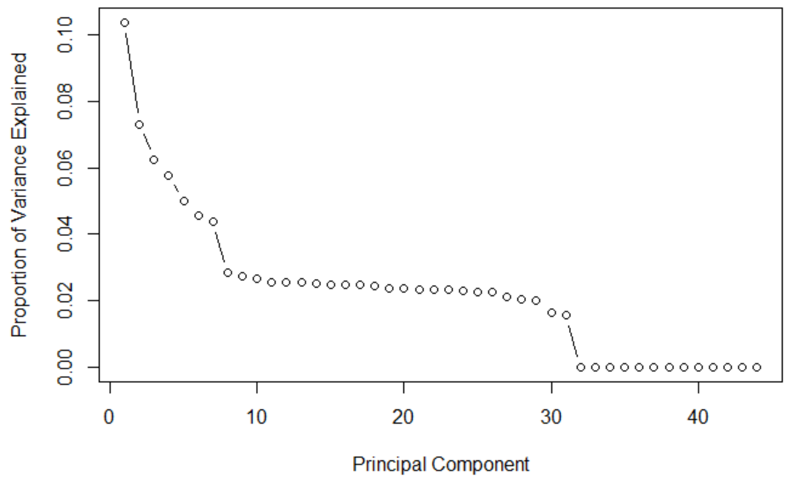
$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_d}$$

고려대학교 KOREA UNIVERSITY

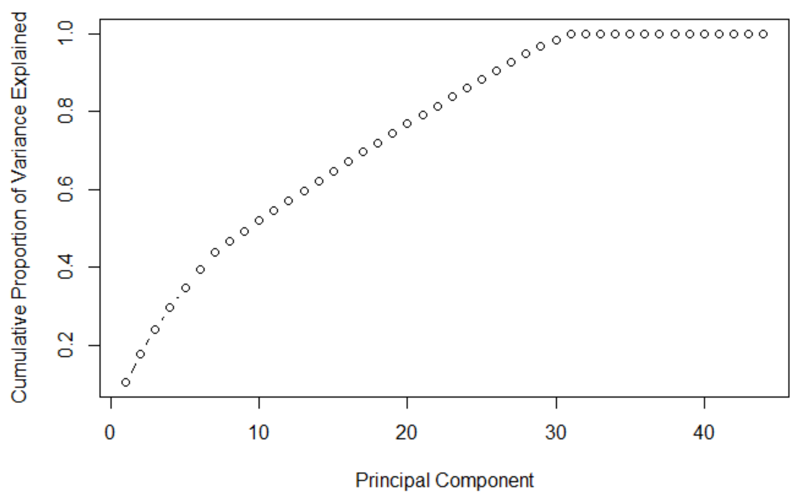
DSBA Data Science & Business Analytics

PCA Issue

- How many principal components are optimal?
  - ✓ Scree plot is commonly used to determine the number of PCs
    - x-axis: index of PCs
    - y-axis: corresponding eigenvalue
  - ✓ Scree plot rapidly decreases in early stages
  - ✓ It does not decrease beyond some points
  - ✓ Selection method 1: Find the elbow point
  - ✓ Selection method 2: Find the smallest PCs that preserve the predetermined required variance



Variance preserved by each PC

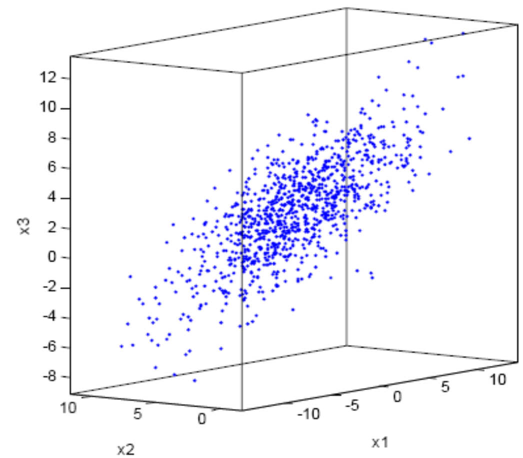


Cumulative variance w.r.t. the number of PCs

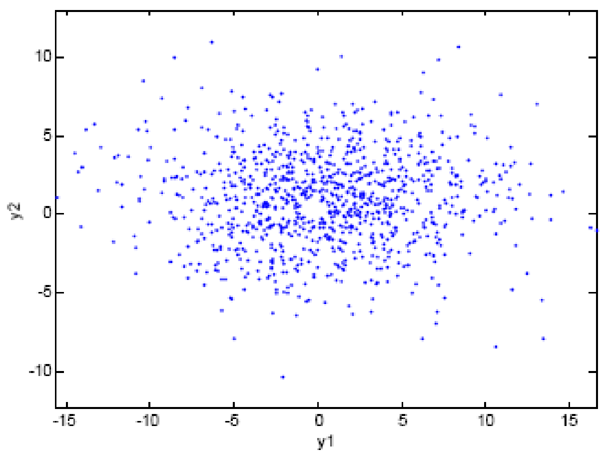
PCA Example 1

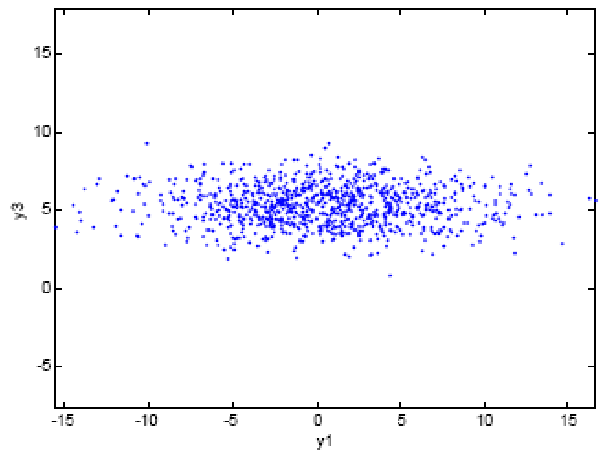
- 3-Dimensional Gaussian distribution
  - ✓ Original means and covariance matrix

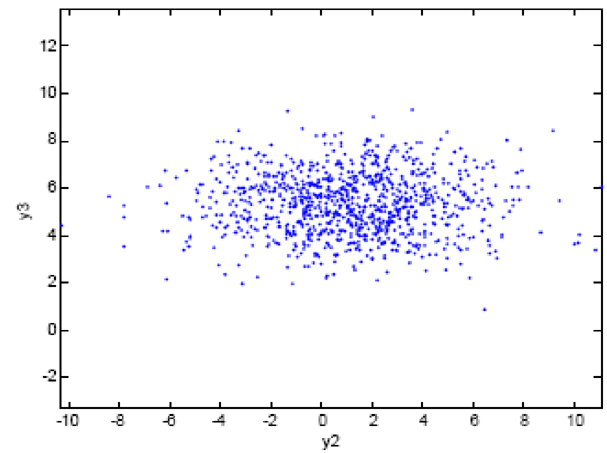
$$\mu = [0\ \ 5\ \ 2]^T \ \text{ and } \ \Sigma = \begin{bmatrix} 25 & -1 & 7 \\ -1 & 4 & -4 \\ 7 & -4 & 10 \end{bmatrix}$$



  - ✓ Projection results for two PCs







PCA Procedure

- PCA Example
  - ✓ Original data

시리얼 이름	제조업체명	유형	칼로리	단백질	지방	나트륨	식이섬유	복합탄수화물	설탕	칼륨	비타민
100% Bran	N	C	70	4	1	130	10	5	6	280	25
100% Natural Bran	Q	C	120	3	5	15	2	8	8	135	0
All-Bran	K	C	70	4	1	260	9	7	5	320	25
All-Bran with Extra Fiber	K	C	50	4	0	140	14	8	0	330	25
Almond Delight	R	C	110	2	2	200	1	14	8		25
Apple Cinnamon Cheerios	G	C	110	2	2	180	1.5	10.5	10	70	25
Apple Jacks	K	C	110	2	0	125	1	11	14	30	25
Basic 4	G	C	130	3	2	210	2	18	8	100	25
Bran Chex	R	C	90	2	1	200	4	15	6	125	25
Bran Flakes	P	C	90	3	0	210	5	13	5	190	25
Cap'n'Crunch	Q	C	120	1	2	220	0	12	12	35	25
Cheerios	G	C	110	6	2	290	2	17	1	105	25
Cinnamon Toast Crunch	G	C	120	1	3	210	0	13	9	45	25
Clusters	G	C	110	3	2	140	2	13	7	105	25
Cocoa Puffs	G	C	110	1	1	180	0	12	13	55	25
Corn Chex	R	C	110	2	0	280	0	22	3	25	25
Corn Flakes	K	C	100	2	0	290	1	21	2	35	25
Corn Pops	K	C	110	1	0	90	1	13	12	20	25
Count Chocula	G	C	110	1	1	180	0	12	13	65	25
Cracklin' Oat Bran	K	C	110	3	3	140	4	10	7	160	25

고려대학교 KOREA UNIVERSITY

DSBA Data Science & Business Analytics

PCA Procedure

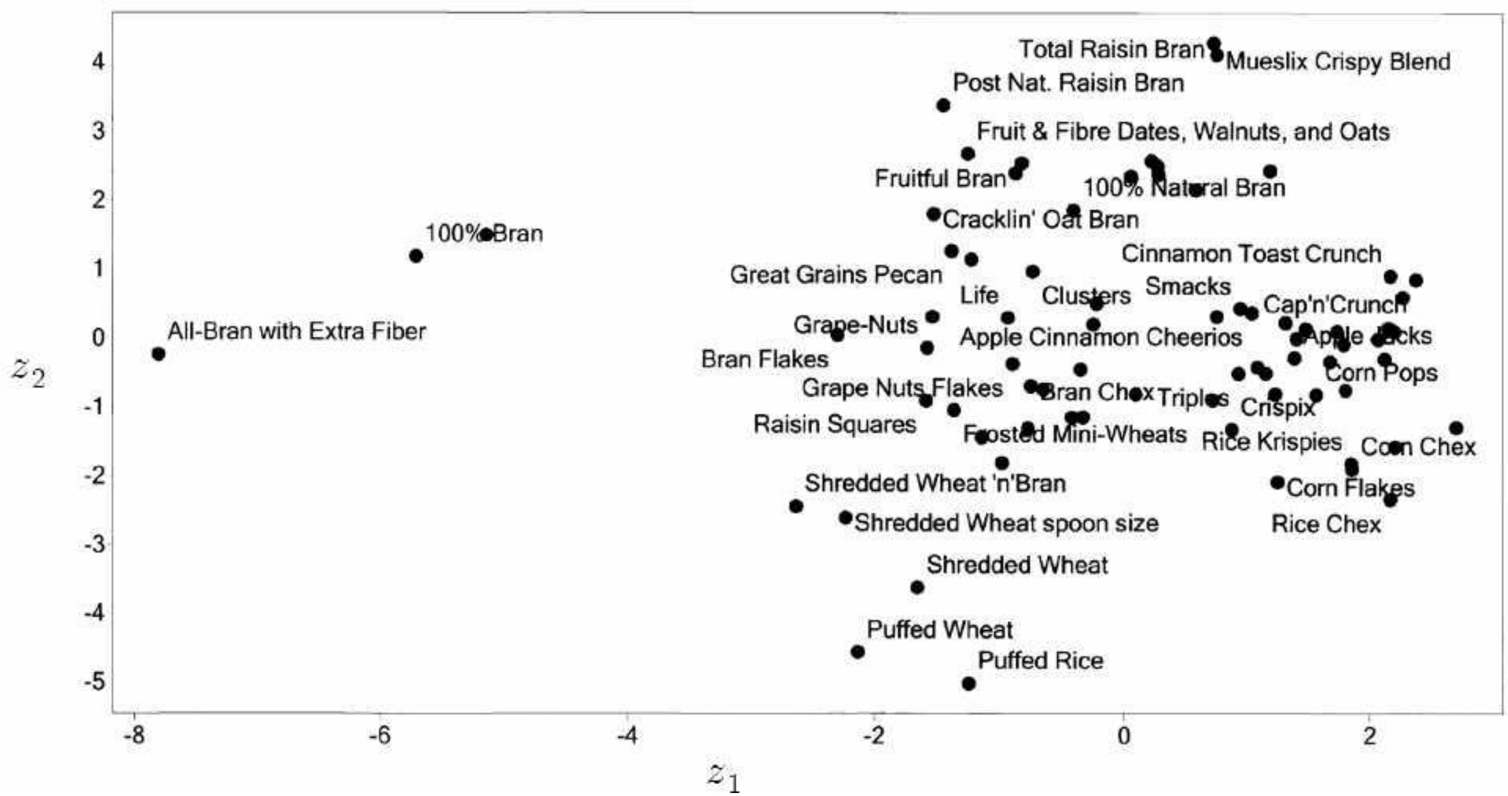
- PCA Example
  - ✓ Eigenvectors eigenvalues for each principal component
    - We need at least 5 PCs to preserve at least 80% of total variance

변수이름	1	2	3	4	5	6	7
calories	0.2995424	0.39314792	0.11485746	0.20435865	0.20389892	-0.25590625	-0.02559552
protein	-0.30735639	0.16532333	0.27728197	0.30074316	0.319749	0.120752	0.28270504
fat	0.03991544	0.34572428	-0.20489009	0.18683317	0.58689332	0.34796733	-0.05115468
sodium	0.18339655	0.13722059	0.38943109	0.12033724	-0.33836424	0.66437215	-0.28370309
fiber	-0.45349041	0.17981192	0.06976604	0.03917367	-0.255119	0.0642436	0.11232537
carbo	0.19244903	-0.14944831	0.56245244	0.0878355	0.18274252	-0.32639283	-0.26046798
sugars	0.22806853	0.35143444	-0.35540518	-0.02270711	-0.31487244	-0.15208226	0.22798519
potass	-0.40196434	0.30054429	0.06762024	0.09087842	-0.14836049	0.02515389	0.14880823
vitamins	0.11598022	0.1729092	0.38785872	-0.6041106	-0.04928682	0.12948574	0.29427618
shelf	-0.17126338	0.26505029	-0.00153102	-0.63887852	0.32910112	-0.05204415	-0.17483434
weight	0.05029929	0.45030847	0.24713831	0.15342878	-0.22128329	-0.39877367	0.01392053
cups	0.29463556	-0.21224795	0.13999969	0.04748911	0.12081645	0.09946091	0.74856687
rating	-0.43837839	-0.25153893	0.1818424	0.0383162	0.05758421	-0.18614525	0.06344455

분산	3.63360572	3.1480546	1.90934956	1.01947618	0.98935974	0.72206175	0.67151642
분산비(%)	27.95081329	24.21580505	14.6873045	7.84212446	7.61045933	5.55432129	5.16551113
누적분산비(%)	27.95081329	52.16661835	66.85391998	74.69604492	82.3065033	87.86082458	93.02633667

PCA Procedure

- PCA Example
  - ✓ In a 2-dimensional space



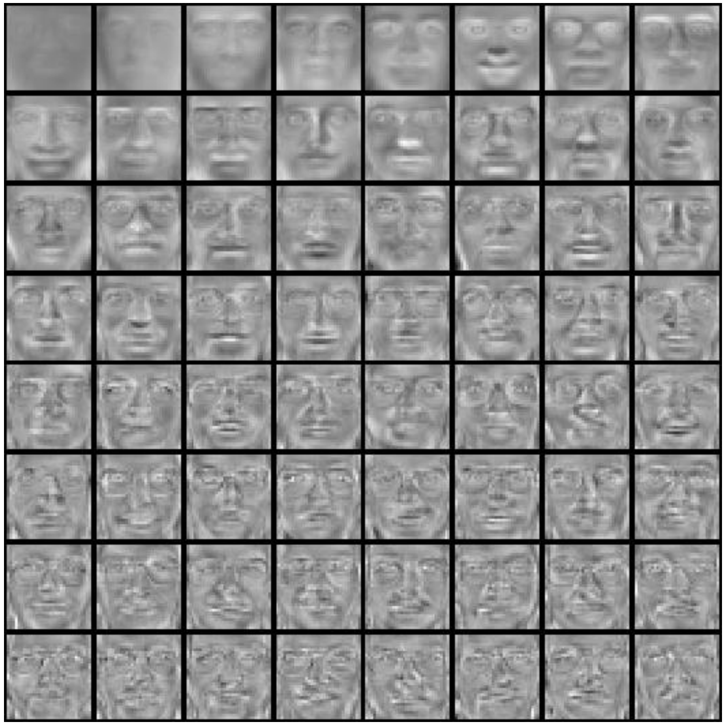
PCA Procedure

- PCA Example
  - ✓ Face image data

Average

Original Image

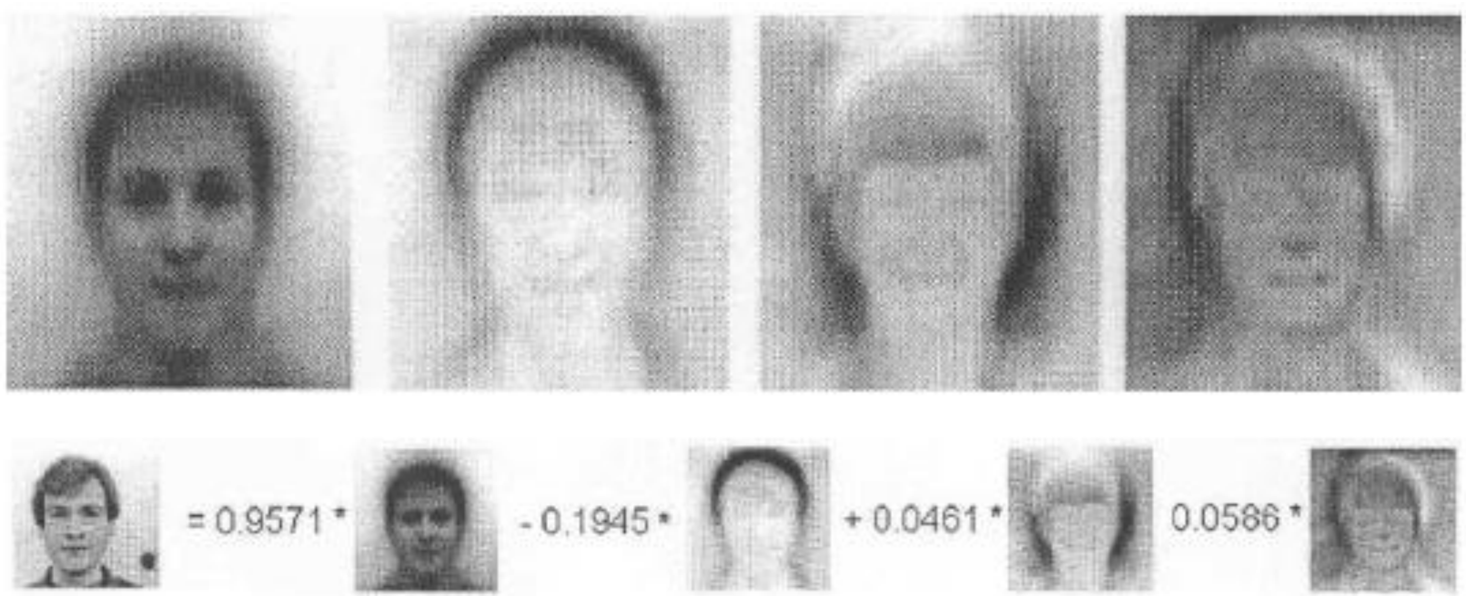
Eigenvectors



PCA Procedure

- PCA Example
  - ✓ Face image data

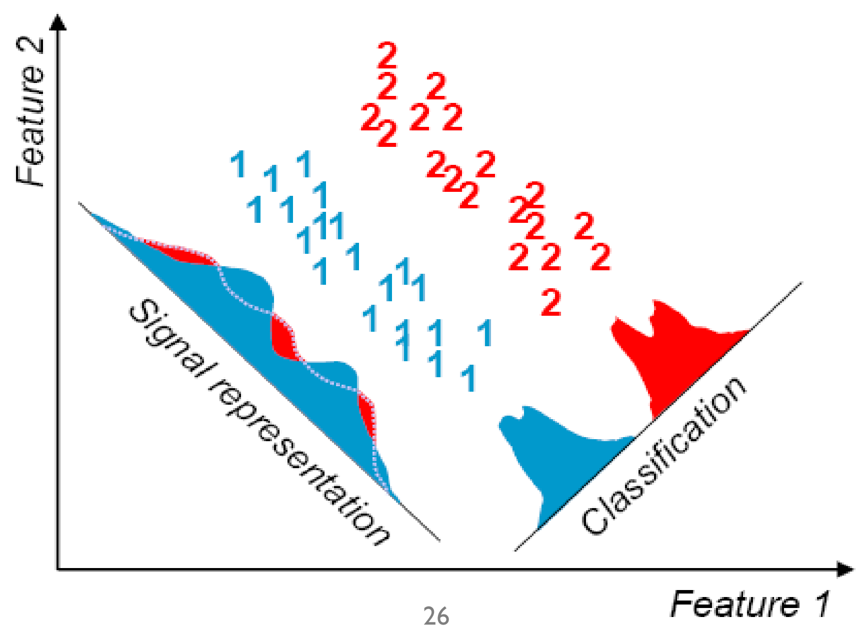
Image reconstruction



PCA: Limitations

- Cannot work well with non-Gaussian or multimodal Gaussian distributions

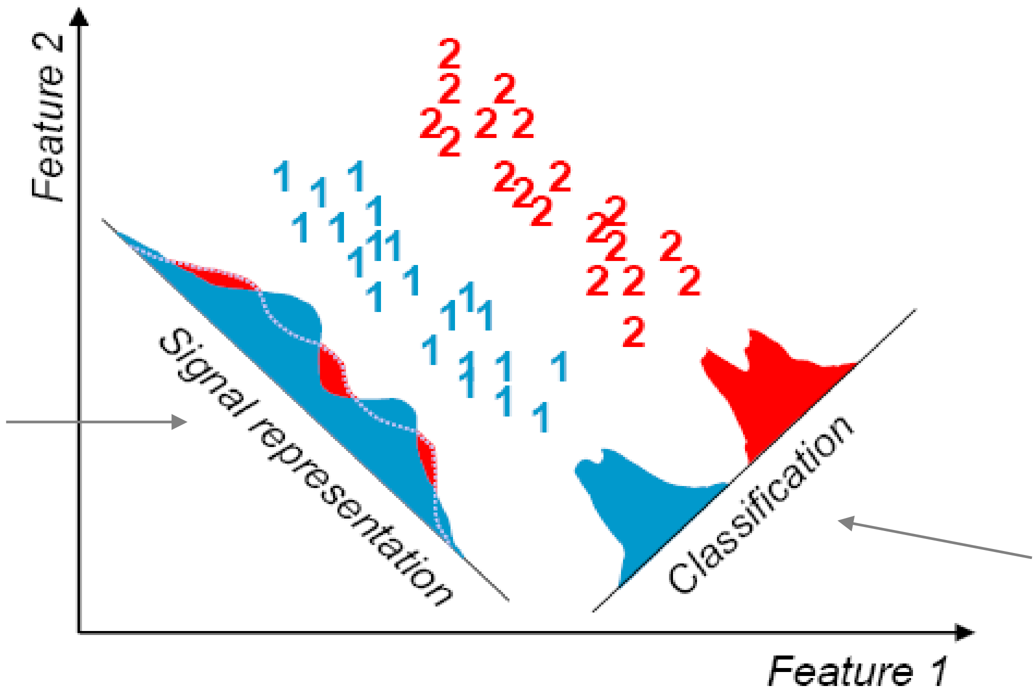
- Not designed for classification



PCA: Limitations

- Not useful for finding the best class decision boundary

Basis found by PCA



Basis found by Fisher's LDA


Q & A

References

<u>Other materials</u>

- Figure in the title page: http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/112-pca-principal-component-analysis-essentials/