Primer on Analysis of Experimental Data and Design of Experiments

Lecture 11. Big Data Classification by Principal Component Analysis

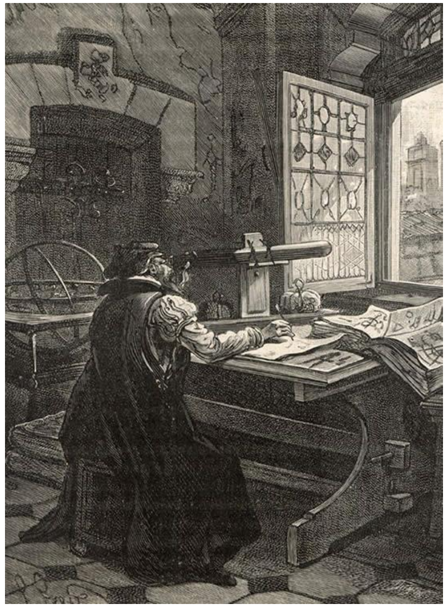
Muhammad A. Alam
alam@purdue.edu

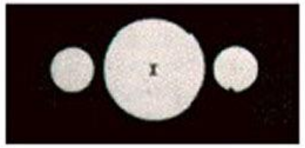
Course Outline

$$\overline{y} = f(\overline{x}) \qquad \overline{x} = x_1, x_2, \ldots x_n \qquad \overline{y} = y_1, y_2, \ldots y_m$$

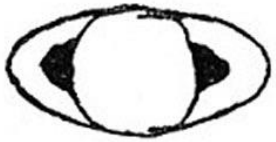
Lecture 1: Introduction
Lecture 2: Collecting and plotting $x_1, x_2, \ldots x_n$
Lecture 3: Physical and empirical $f, F, df/dx, \ldots$
Lecture 4: Model selection between $f_1, f_2, \ldots$
Lecture 5: Model Selection: Cross-validation and Bootstrapping method
Lecture 6: Scaling theory with known f, $f(\overline{x}) = f(\overline{X})$
Lecture 7: Scaling theory with unknown f, $\overline{x} \to X$
Lecture 8: Design of experiments to determine $\overline{y}_{\max} = f(\overline{x})$
Lecture 9: DOE and ANOVA
Lecture 11: Principle component analysis for classifying $\{y\}$.
Lecture 12: Machine learning … Statistical approach to learn f
Lecture 13: Machine Learning ….Additional Concepts
Lecture 14: Interpretable ML: Physics-based machine learning and system equation modeling
Lecture 14: Conclusions

Small vs. big data

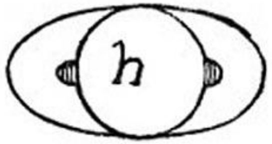




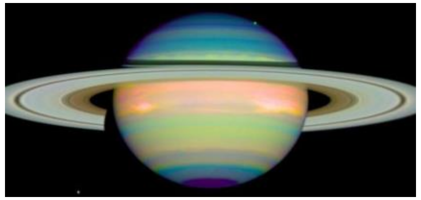
Galileo first sketch
1610



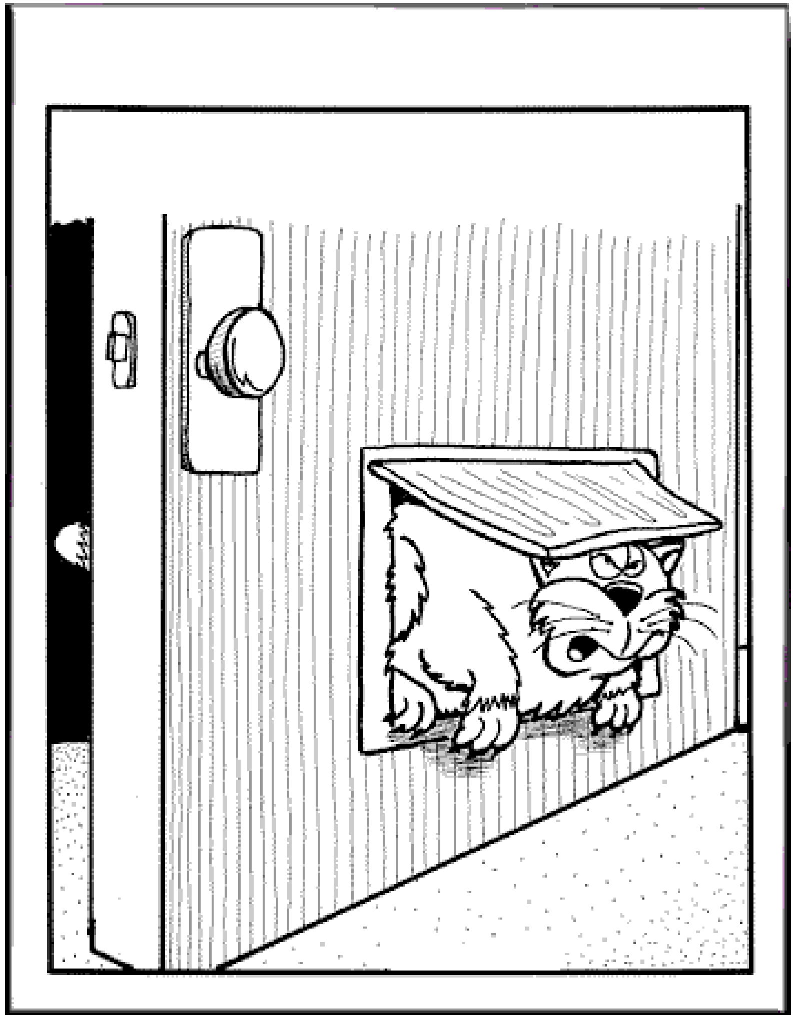
Better telescope
1616



Published etch
1623



“Big data” techniques apply to “little data” too



Isaac Newton, known as a physicist, mathematician and astronomer, may have also been the “cat door inventor”!According to an anecdote, Newton foolishly made a large hole for the mother cat and six small holes for her six kittens, not understanding that the kittens could follow their mother through the large hole!

Analysis of big data

Data

Data Processing

Distance Metrics(2.1)

Sampling (2.2)

Dimensionality Reduction (2.3);

PCA (2.3.1) and

SVD (2.3.2)

Model Leanring

Supervised

Regression, Classifications

kNN (3.1.1), Decision Trees (3.1.2), Rules (3.1.3), Bayesian Classification (3.1.4), Logistic Regression (3.1.5), SVM (3.1.5), and ANN (3.1.7)

Unsupervised Learning

Associated Rules, Matrix Completion, Clustering)

k-means (4.1.1), Density-based (4.1.2), Message passing(4.1.2), Hierarchical (4.1.2), LDA (4.1.2), Bayesian Non-parametric (4.1.2), LSH (4.1.2)

Testing and Validation

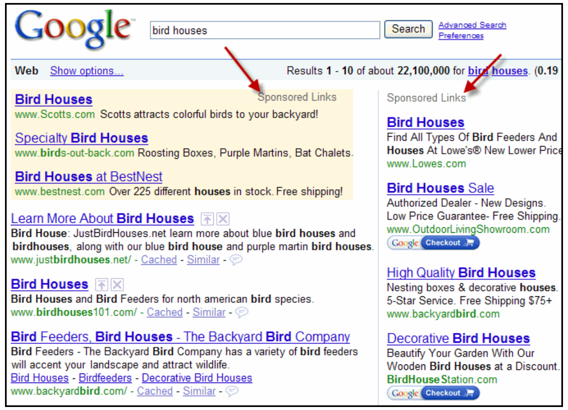
Evaluating classifiers (3.3)

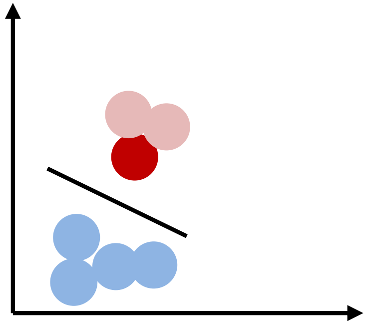
Outline

1. Introduction
2. Why do we need reduction in data dimension
2. Theory of Principle Component Analysis
3. Applications of Principle Component Analysis
4. Conclusions

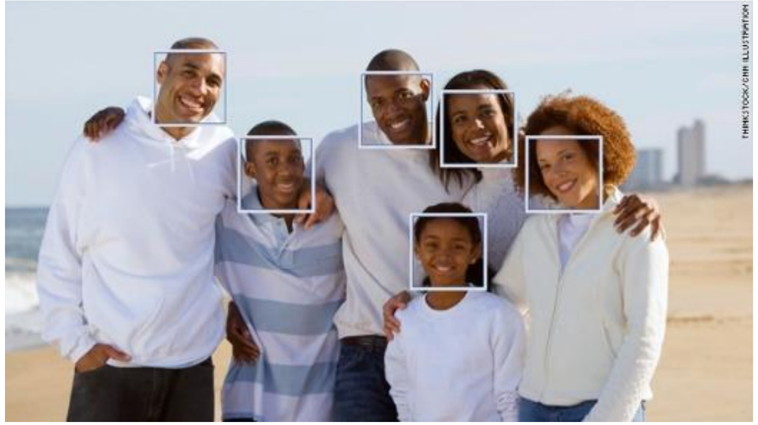
Classification problem in big data

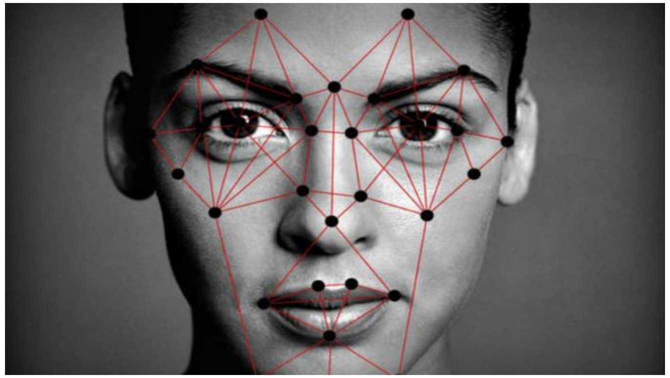
Advertisement
Recommendation



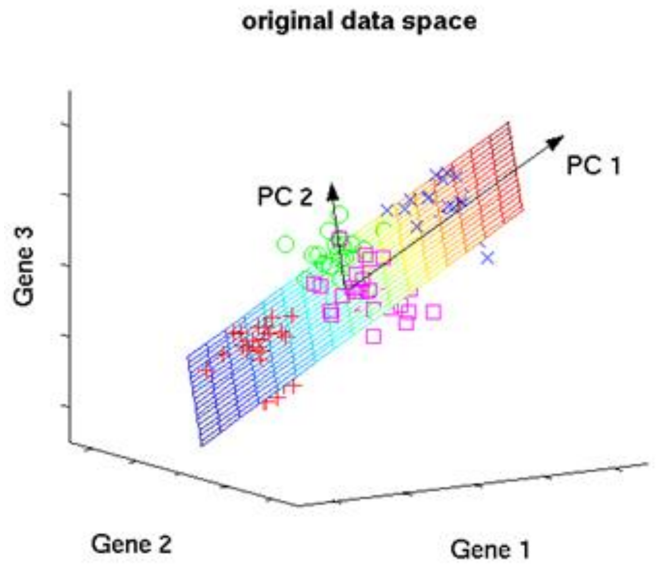


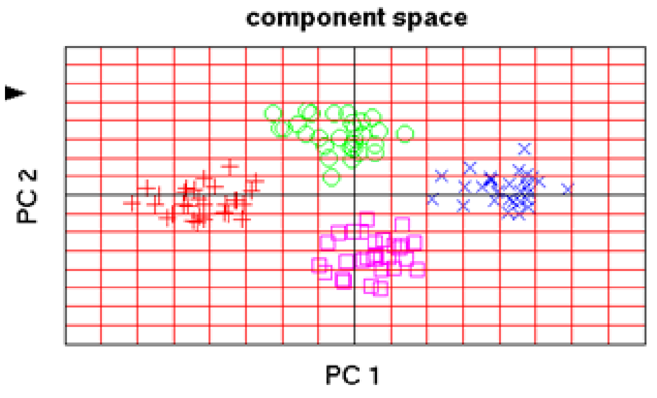
Facial Recognition
Voice Recognition
Spam Filtering





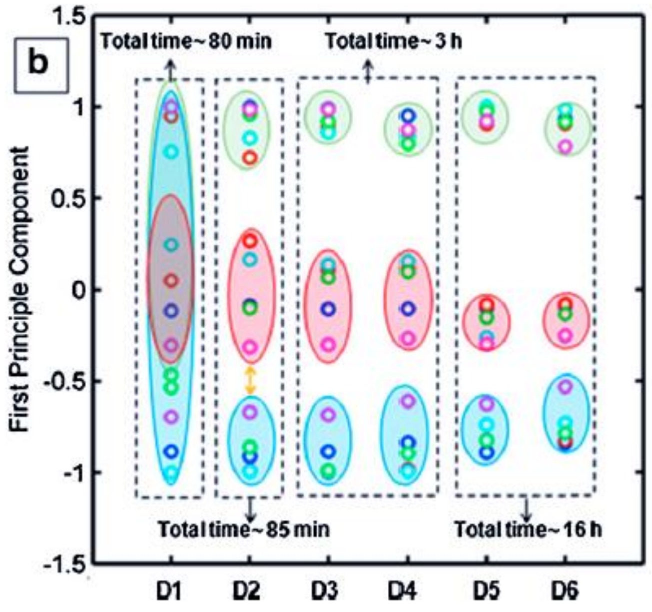
PCA helps classification



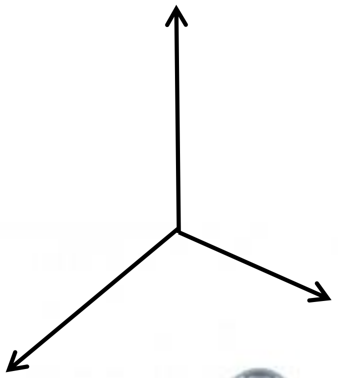


name	100m	Long.jump	//	Javeline	1500m	Rank	Points	Competition
SEBRLE	11.04	7.58		63.19	291.7	1	8217	Decastar
CLAY	10.76	7.4		60.15	301.5	2	8122	Decastar
Macey	10.89	7.47		58.46	265.42	4	8414	OlympicG
Warners	10.62	7.74		55.39	278.05	5	8343	OlympicG
\\								
Zsivoczky	10.91	7.14		63.45	269.54	6	8287	OlympicG
Hernu	10.97	7.19		57.76	264.35	7	8237	OlympicG
Pogorelov	10.95	7.31		53.45	287.63	11	8084	OlympicG
Schoenbeck	10.9	7.3		60.89	278.82	12	8077	OlympicG
Barras	11.14	6.99		64.55	267.09	13	8067	OlympicG
KARPOV	11.02	7.3		50.31	300.2	3	8099	Decastar
WARNERS	11.11	7.6		51.77	278.1	6	8030	Decastar
Nool	10.8	7.53		61.33	276.33	8	8235	OlympicG
Drews	10.87	7.38		51.53	274.21	19	7926	OlympicG

Active individuals
Active variables
Supplementary **quantitative** variables
Supplementary **qualitative** variable
Supplementary individuals



PCA Also help in data compression



3D information projected onto a 2D plane

Perspective projection invented by Fillipo Brunelleschi & Masaccio

Outline

1. Why do we need reduction in data dimension

2. Theory of Principle Component Analysis

3. Applications of Principle Component Analysis

4. Conclusions

Principle Component Analysis (PCA)

- Example: Are some students falling behind? Difficult to decide in a multidimensional data

N samples

$\overbrace{\text{Math} \quad \text{English}}^{\text{2 measurements}}$

Student	Math	English
1	97	94
2	32	60
3	89	76
4	50	45
…		
N	90	86

1 measurement

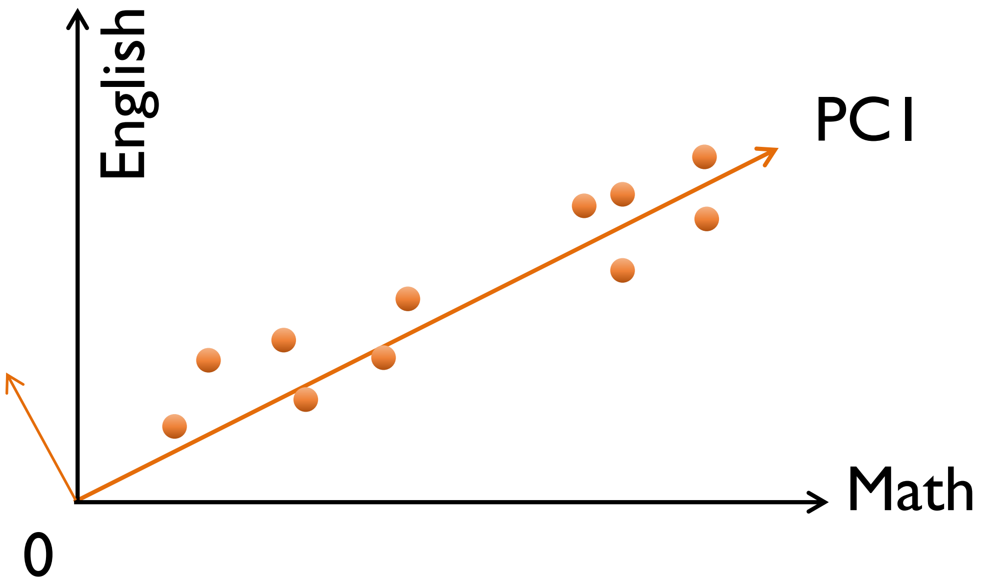
Student	PC1
1	
2	
3	
4	
…	
N	

Basic Concept of PCA

2 measurements

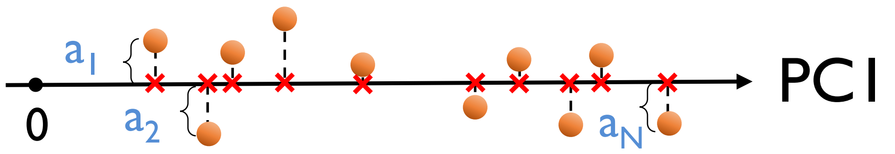
N samples

Student	Math	English
1	97	94
2	32	60
3	89	76
4	50	45
…		
N	90	86



Rotate & redefine the new axis: PC1

To reduce 2-D data to 1-D data: find a direction onto which to project the data so as to *minimize the projection error*



Projection error $= \sum_N a_N^2$

PCA through Singular Value Decomposition

2 measurements

N samples

Student	Math	English
1	97	94
2	32	60
3	89	76
4	50	45
…		
N	90	86

$$\longrightarrow \quad X = U\Sigma V^{\mathrm{T}}$$

Weight of each principal components

Direction of the new axis (PC1, PC2 …)

m measurements

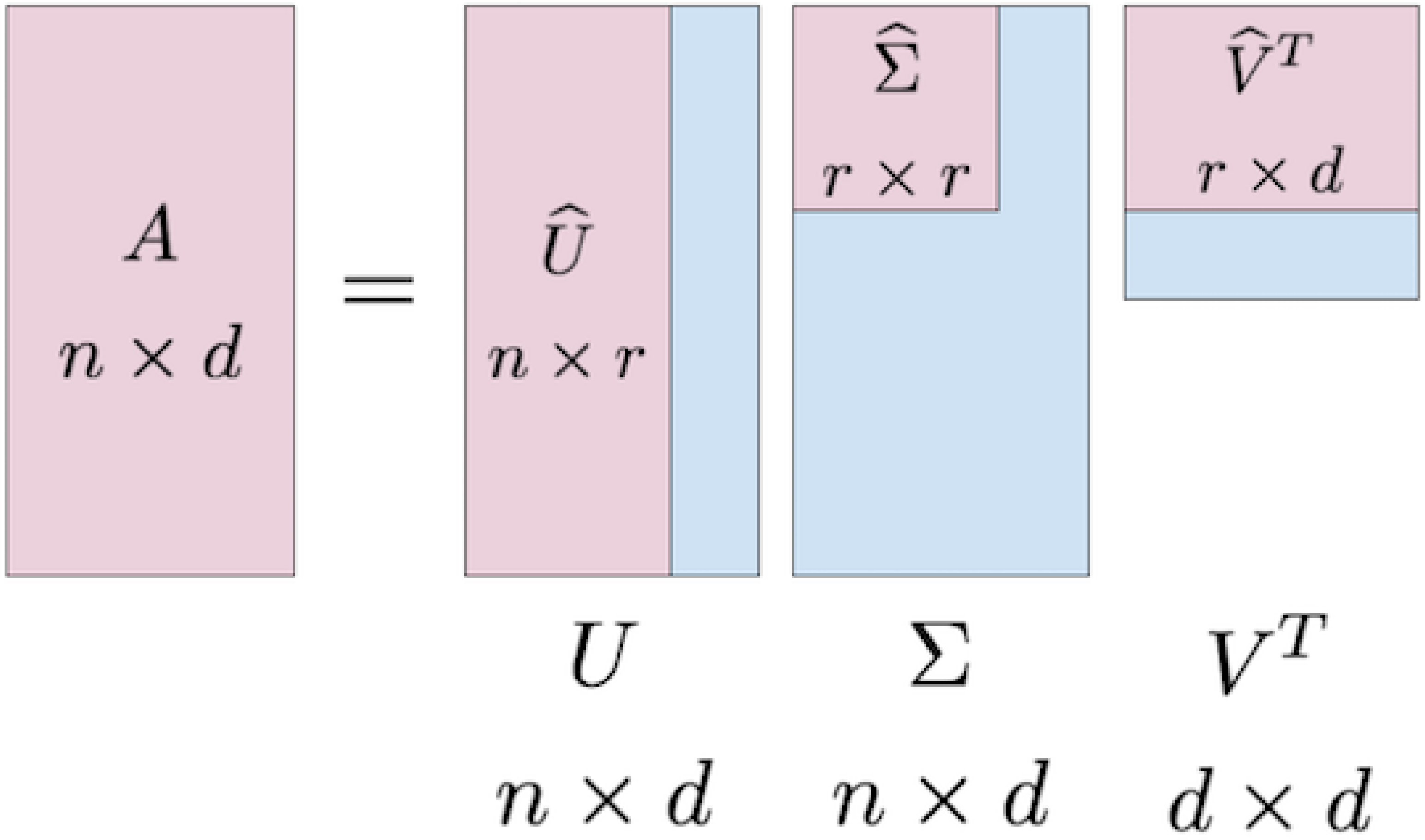
$$\underbrace{\begin{bmatrix} \boldsymbol{X}: \\ n \times m \\ \text{matrix} \end{bmatrix}}_{n \text{ samples}} = \begin{bmatrix} u_{11} & u_{21} & \cdots & u_{n1} \\ u_{12} & u_{22} & & u_{n2} \\ \cdots & \cdots & & \\ u_{1n} & u_{2n} & \cdots & u_{nn} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & \\ \cdots & & \cdots & \\ 0 & 0 & & \sigma_m \end{bmatrix} \begin{bmatrix} v_{11} & v_{21} & \cdots & v_{m1} \\ v_{12} & v_{22} & & v_{m2} \\ \cdots & \cdots & & \\ v_{1m} & v_{2m} & \cdots & v_{mm} \end{bmatrix}$$

$\boldsymbol{U}$: $n \times n$ matrix

$\boldsymbol{\Sigma}$: $n \times m$ diagnal matrix

$\boldsymbol{V}$: $m \times m$ matrix

Reduce dimension by Singular Value Decomposition



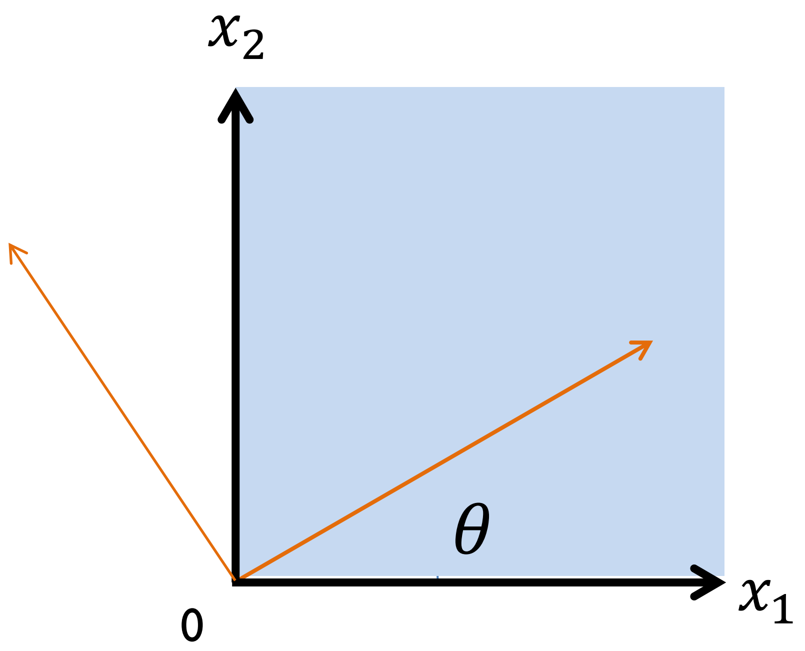
Example 1: Rotation matrix

$$R = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

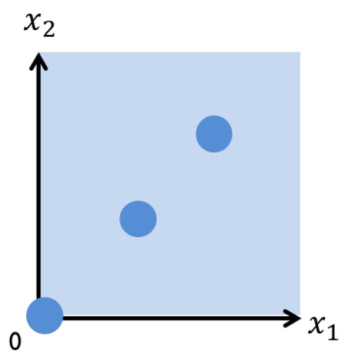
$$R^{90} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$$R^{180} = \begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix}$$

$$R^{270} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$



SVD rotates the axes optimally



(3x2 matrix)

$$X = \begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 0 & 0 \end{pmatrix}$$

Three points (1,1), (2,2) and (0,0)

SVD ({{1,1},{2,2},{0,0}})

$$X \;=\; U\,\Sigma\,V^T$$

(2x2 matrix)

$$V^T = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

Rotate by 45 degrees

The PC is sufficient

$$U = \begin{pmatrix} 1/\sqrt{5} & 0 & -2/\sqrt{5} \\ 2/\sqrt{5} & 0 & 1/\sqrt{5} \\ 0 & 1 & 0 \end{pmatrix}$$

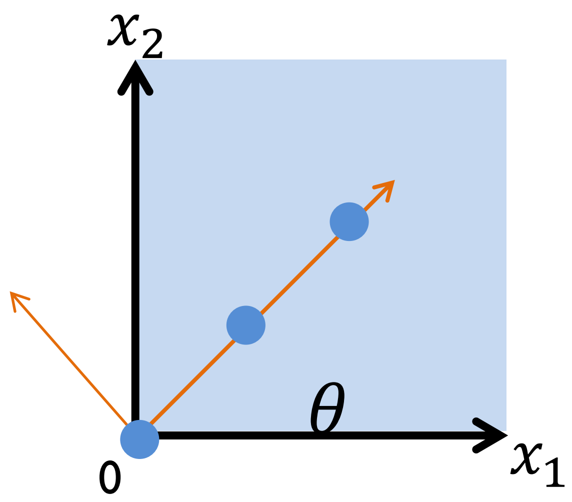
(3x3 matrix)

(3x2 matrix)

$$\Sigma = \begin{pmatrix} \sqrt{10} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

SVD components allows reconstruction

$$X = \begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 0 & 0 \end{pmatrix} = u_1 \, \sigma_1 \, v_1^T + u_2 \, \sigma_2 \, v_2^T$$

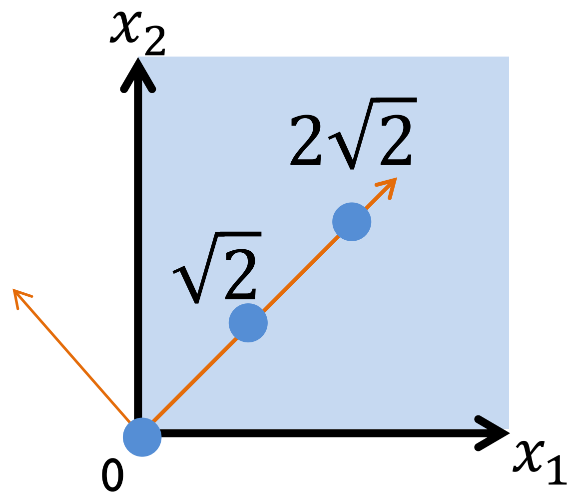


$$u_1 = \begin{pmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \\ 0 \end{pmatrix} \qquad \sigma_1 = \sqrt{10} \qquad v_1^T = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 0 & 0 \end{pmatrix} \qquad X' = u_1 \, \sigma_1 \, v_1^T = \begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 0 & 0 \end{pmatrix}$$

$X' = X$ because the projection is exact

Projection along PCs



$$XV = U\Sigma V^T V = U\Sigma$$

(3x2 matrix)

$$X = \begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 0 & 0 \end{pmatrix}$$

(2x2 matrix)

$$V = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \rightarrow X.V = \begin{pmatrix} \sqrt{2} & 0 \\ 2\sqrt{2} & 0 \\ 0 & 0 \end{pmatrix}$$

$$U = \begin{pmatrix} 1/\sqrt{5} & 0 & -2/\sqrt{5} \\ 2/\sqrt{5} & 0 & 1/\sqrt{5} \\ 0 & 1 & 0 \end{pmatrix}$$

(3x3 matrix)

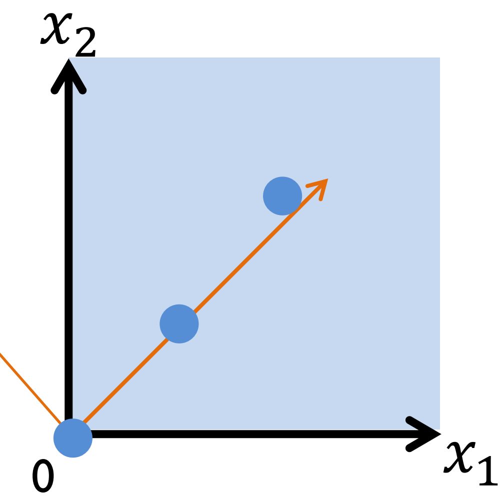
$$\Sigma = \begin{pmatrix} \sqrt{10} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \rightarrow U.\Sigma = \begin{pmatrix} \sqrt{2} & 0 \\ 2\sqrt{2} & 0 \\ 0 & 0 \end{pmatrix}$$

(3x2 matrix)

(3x2 matrix)

Projection along PC1

Example 2: More general result



$$X = \begin{pmatrix} 0 & 0 \\ 1 & 1 \\ 2 & 2.1 \end{pmatrix}$$

(3x2 matrix)

Three points (0,0), (1,1), (2,2.1)

SVD ({0,0}, {1,1},{2,2.1})

(2x2 matrix)

$$V^T = \begin{pmatrix} 0.693 & 0.721 \\ 0.721 & -0.693 \end{pmatrix}$$

$$X = U \, \Sigma \, V^T$$

$$U = \begin{pmatrix} \sim 0 & \sim 0 & 1 \\ 0.438 & 0.899 & \sim 0 \\ 0.899 & -0.438 & 0 \end{pmatrix}$$

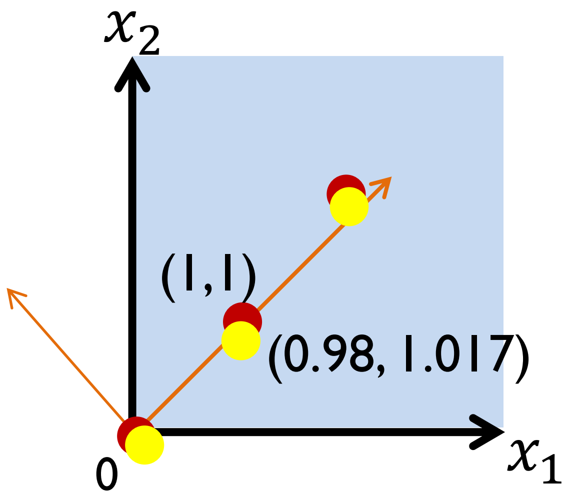
(3x3 matrix)

$$\Sigma = \begin{pmatrix} 3.226 & 0 \\ 0 & 0.031 \\ 0 & 0 \end{pmatrix}$$

(3x2 matrix)

SVD approximates the exact result

$$X = \begin{pmatrix} 0 & 0 \\ 1 & 1 \\ 2 & 2.1 \end{pmatrix} = u_1\,\sigma_1\,v_1^T + u_2\,\sigma_2\,v_2^T$$



$$u_1 = \begin{pmatrix} 0 \\ -0.438 \\ -0.721 \end{pmatrix} \quad \sigma_1 = 3.226 \quad v_1^T = \begin{pmatrix} -0.693 & -.721 \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 0 & 0 \end{pmatrix} \qquad X' = u_1\,\sigma_1\,v_1^T = \begin{pmatrix} 0 & 0 \\ 0.98 & 1.017 \\ 2.01 & 2.08 \end{pmatrix}$$

$X' \sim X$ because the projection is approximate

(continued) Projection along PCs

$$XV = U\Sigma V^T V = U\Sigma$$

(3x2 matrix)

$$X = \begin{pmatrix} 0 & 0 \\ 1 & 1 \\ 2 & 2.1 \end{pmatrix}$$

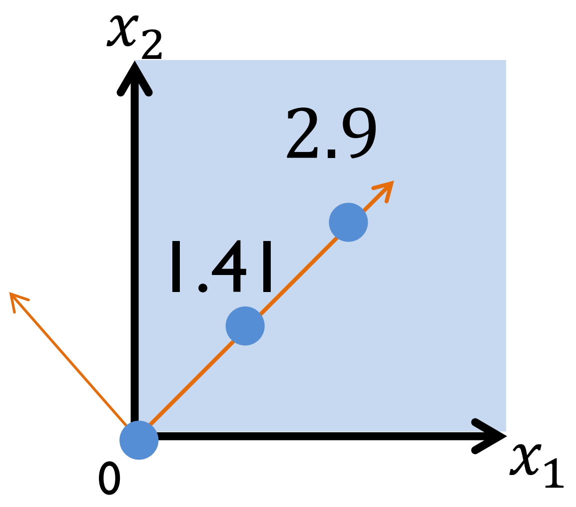
(2x2 matrix)

$$V = \begin{pmatrix} -0.693 & -0.721 \\ -0.721 & 0.693 \end{pmatrix}$$

(3x2 matrix)

$$X.V = U.\Sigma = \begin{pmatrix} 0 & 0 \\ 1.414 & -0.028 \\ 2.9 & 0.0133 \end{pmatrix}$$

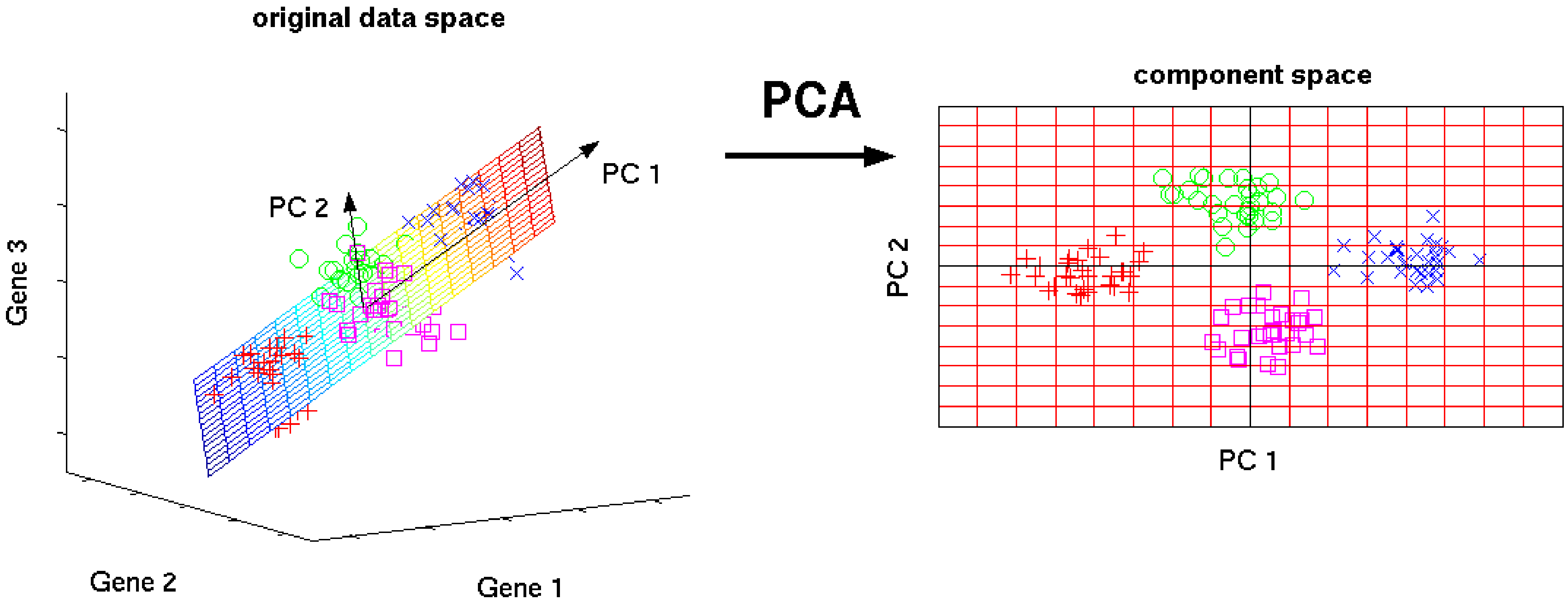
Projection along PC1



Outline

1. Why do we need reduction in data dimension
2. Theory of Principle Component Analysis
3. Applications of Principle Component Analysis
4. Conclusions

Principle Component Analysis for classification



If you like this book, you will also like that book (because you belong to the same category)

Image Transmission by Principle Component Analysis

$$X_{1000\times500} = \underbrace{u_1}_{1000\text{x}1}\ \overset{1\text{x}1}{\sigma_1}\ \underbrace{v_1^T}_{1\text{x}500} + u_2\ \sigma_2\ v_2^T + \ \dots.$$

MATLAB code (by Camsari)

```
%% SVD - Image processing
clearvars;clc;close all;
%%

IMM=imread('MonaLisa.jpg');
IMM=im2double(IMM);
imshow(IMM);
%% Turn it into 2D (grayscale)
IM2D=rgb2gray(IMM);
%% Original in Grayscale
figure(123)
subplot(2,2,1)
imshow(IM2D);
title('Original')
```

```
%% SVD Decomposition
[U,S,V]=svd(IM2D);
%% Optional:Plot the diagonals
%figure(234)
%semilogy(diag(S))

%% Keep the first 50 dimensions.
NR = 50;AR50=zeros(size(IM2D));
for ii=1:NR
  AR50=AR50+U(:,ii)*S(ii,ii)*V(:,ii)';
end

%% Keep the first 15 dimensions.
NR = 15;AR15=zeros(size(IM2D));
for ii=1:NR
  AR15=AR15+U(:,ii)*S(ii,ii)*V(:,ii)';
end
```

Conclusions

1. PCR is a powerful tool to classify multi-dimensional data (e.g. postal codes in handwritten envelops)

2. PCR decomposition by SVD provides both the rotated axes (V) and the projection on the rotated axes ($U\Sigma$). Each column of ($U\ \Sigma$) is the projection on that principal component.

3. If there are 100 dimensions and 5 key distinguishing features, then top five singular values may not align with the top five features. One should keep approximately 20 to preserve the top 5 features.

4. The desired by accuracy is obtained by choosing a k such that $p = r_k = \sum_{i=1}^{k} \lambda_i^2 / \sum_{i=1}^{N} \lambda_i^2$.

5. Other techniques (e.g. Fisher linear discriminators) which finds the direction of the line that best separates two classes may be more accurate or efficient. For example, in Facial recognition, the PCA eigenvalues are called eigenfaces, while that from Fisher LDA is called Fisher's faces.

Review Questions

1. What is “singular” about singular value decomposition?
2. What is the physical meaning of U and V?
3. How many Principal Components should we need to keep? How do you quantify it ?
4. What are the disadvantages of SVD-based classification? In what ways is machine-learning better?
5. What other methods of classification do we have?
6. What applications do we have SVD other than classification (e.g. data compression, etc.)?
7. Taken from your daily experience, Give several examples where SVD classification can be useful.
8. Can you do SVD with Excel? What about Wolfram alpha?

References

Principal Component Analysis

A tutorial on Principle Component Analysis, J. Shlens, arxib 2009. (shlens@salk.edu)

For an interesting application in PCA, see "Recommended for you", J.A. Konstan and J. Riedl, IEEE Spectrum, p. 55, Oc. 2012.

Limitation of PCR:
http://www.svcl.ucsd.edu/courses/ece271B-F09/handouts/Dimensionality2.pdf

Athlete PCR example is taken from:
http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/112-pca-principal-component-analysis-essentials

Strang, Gilbert, et al. Introduction to linear algebra. Vol. 3. Wellesley, MA: Wellesley-Cambridge Press, 1993. provides an excellent introduction to PCR and SVD methods.

Linear Discriminant Analysis (LDA) – A very good numerical example is posted here.
http://people.revoledu.com/kardi/tutorial/LDA/Numerical%20Example.html

Bacteria Osmoregulation Example is taken from Ebrahimi, Aida, and Muhammad A. Alam. "Time-resolved PCA of 'droplet impedance' identifies DNA hybridization at nM concentration." Sensors and Actuators B: Chemical 215 (2015): 215-224.

A wonderful set of lectures by Stuart Hunter is available in youtube, see …
http://www.youtube.com/watch?v=AVUAt0Qly60&list=PLWQ-BDMTHPQVH3IUGF7EM_3XHJWFD2EIP