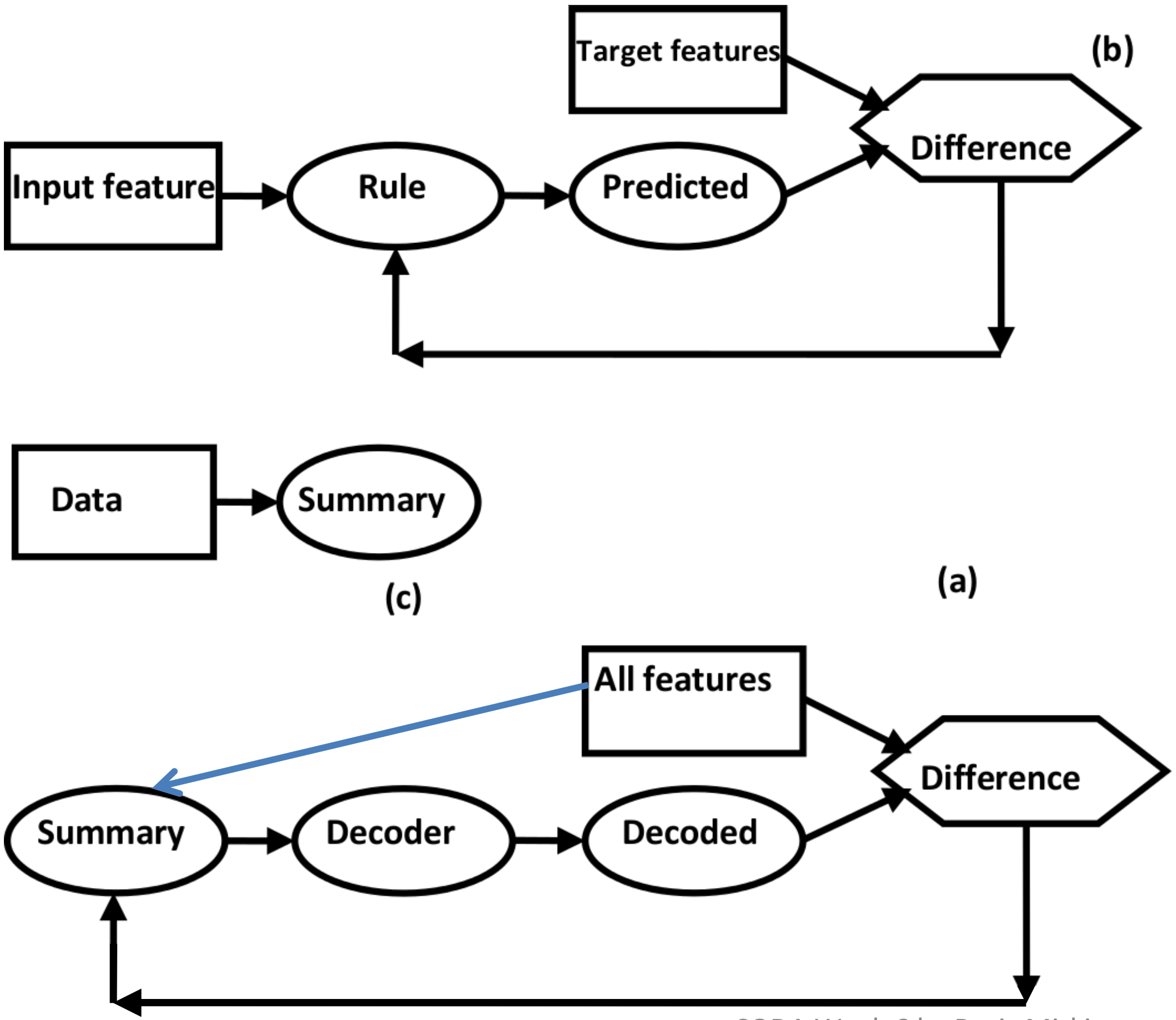
Week 6. Principal Component Analysis: Method and Model

CONTENTS

Week 6. Principal Component Analysis: P.1 Summarization versus Correlation

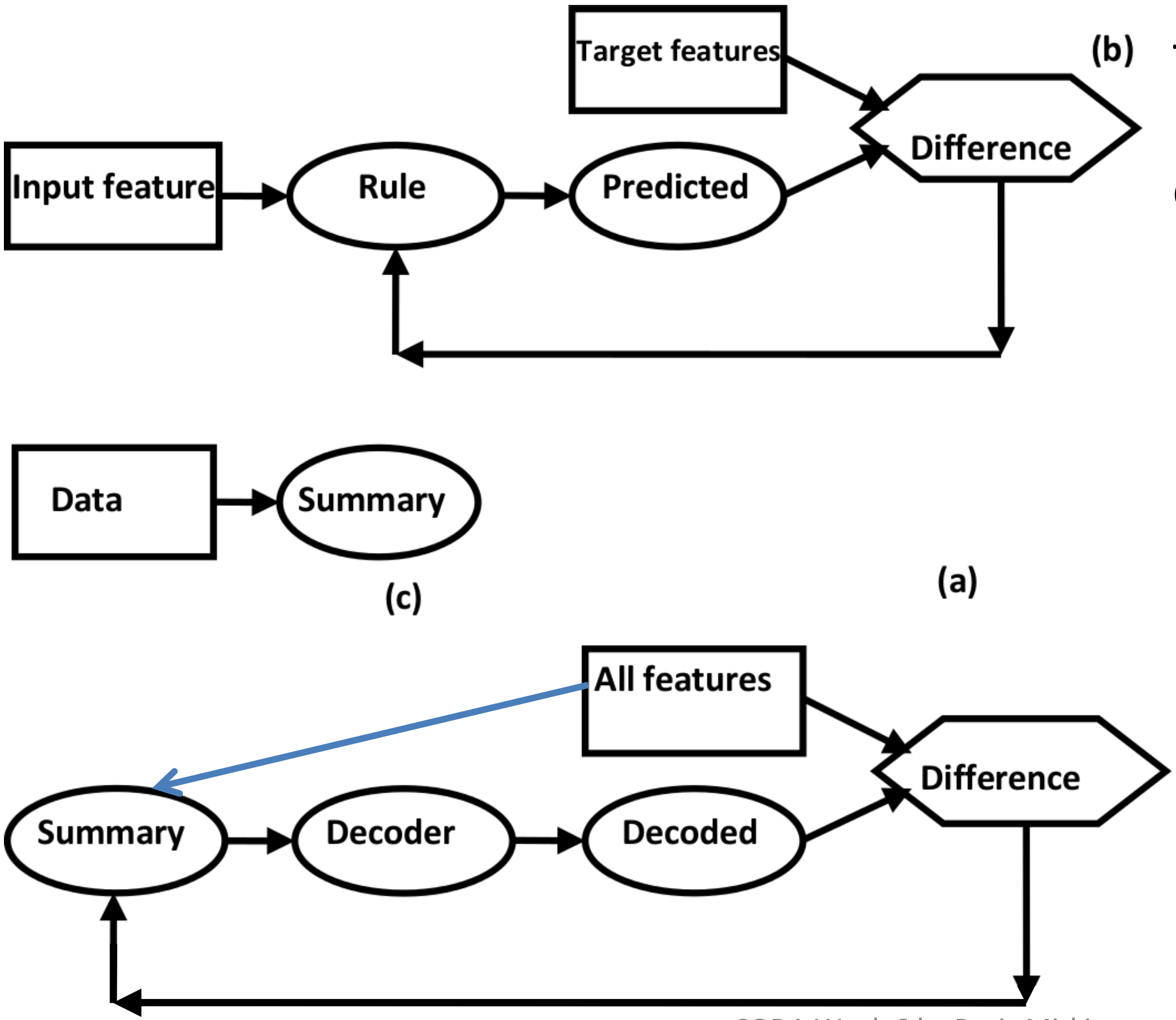


(b) Correlation problem: data recovery formulation

(c) Conventional view of summarization

(a) Data recovery view of summarization (all features are target)

Week 6. Principal Component Analysis
P.1 Issue of data standardization



(b) Correlation problem: target feature scaling predetermines the scale of the output

(a) Summarization outcome much depends on data preprocessing because the recovery extent is measured by the sum of deviations. Deviations change at changing the scale.

Week 6. Principal Component Analysis: P.1 Data standardization issue

Example: Consider Company data

Company name	Income, $mln	MShare,%	NSup	EC	Sector
Aversiona	19.0	43.7	2	No	Utility
Antyops	29.4	36.0	3	No	Utility
Astonite	23.9	38.0	3	No	Industrial
Bayermart	18.4	27.9	2	Yes	Utility
Breaktops	25.7	22.3	3	Yes	Industrial
Bumchista	12.1	16.9	2	Yes	Industrial
Civiok	23.9	30.2	4	Yes	Retail
Cyberdam	27.2	58.0	5	Yes	Retail

Company data 8×5 **converted to the quantitative** format 8×7

Entity	Income	MSchar	NSup	EC?	Util?	Indu?	Retail?
1	19.0	43.7	2	0	1	0	0
2	29.4	36.0	3	0	1	0	0
3	23.9	38.0	3	0	0	1	0
4	18.4	27.9	2	1	1	0	0
5	25.7	22.3	3	1	0	1	0
6	12.1	16.9	2	1	0	1	0
7	23.9	30.2	4	1	0	0	1
8	27.2	58.0	5	1	0	0	1

Week 6. Principal Component Analysis: P.1 Data standardization issue, 1

Company data 8×5 converted to the quantitative format 8×7

#	Inco	MSch	NSup	EC	Util	Indu	Reta
1	19.0	43.7	2	0	**1**	**0**	**0**
2	29.4	36.0	3	0	**1**	**0**	**0**
3	23.9	38.0	3	0	**0**	**1**	**0**
4	18.4	27.9	2	1	**1**	**0**	**0**
5	25.7	22.3	3	1	**0**	**1**	**0**
6	12.1	16.9	2	1	**0**	**1**	**0**
7	23.9	30.2	4	1	**0**	**0**	**1**
8	27.2	58.0	5	1	**0**	**0**	**1**

Three standardizations:

(i) Just centering by subtraction within column means

(ii) Centering and normalization by dividing over column ranges

(iii) Centering and normalization as in (ii) plus additionally dividing the three Sector columns on the right by $\sqrt{3}$ to compensate for the multiple columns

Week 6. Principal Component Analysis: Data standardization issue, 2

Company data 8×5 converted to the quantitative format 8×7

#	Inco	MSch	NSup	EC	Util	Indu	Reta
1	19.0	43.7	2	0	1	0	0
2	29.4	36.0	3	0	1	0	0
3	23.9	38.0	3	0	0	1	0
4	18.4	27.9	2	1	1	0	0
5	25.7	22.3	3	1	0	1	0
6	12.1	16.9	2	1	0	1	0
7	23.9	30.2	4	1	0	0	1
8	27.2	58.0	5	1	0	0	1

Three standardizations:

(i). (ii) and (iii)

Why are that many, and what is the need in data standardization?

Goal: to sharpen the data structure

Data standardization:

A. **Feature centering: to look at feature values against a "normal" backdrop**

B. **Feature normalization: to balance feature weights**

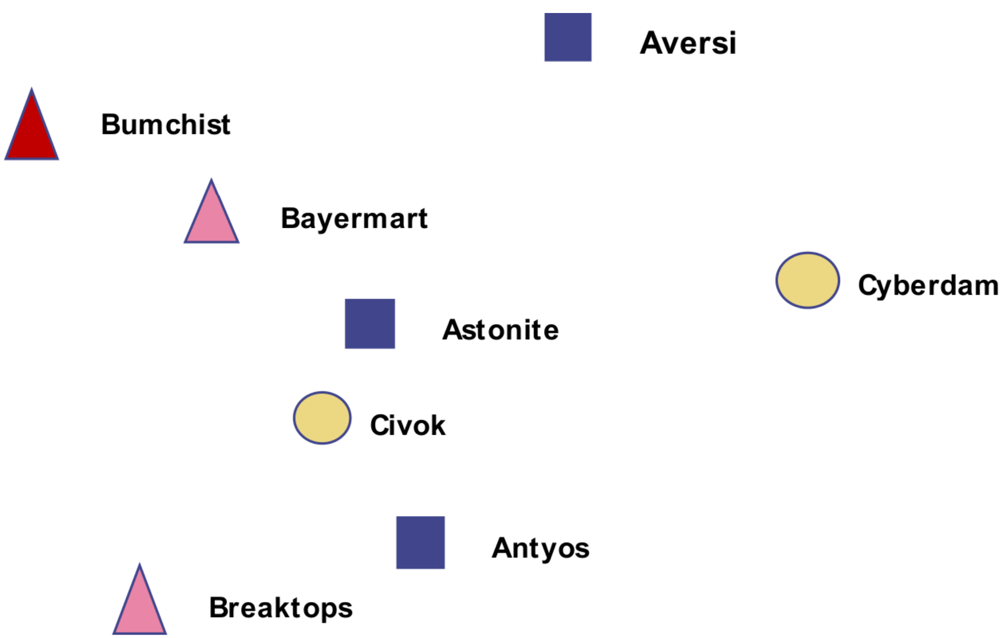
Week 6. Principal Component Analysis: Data standardization issue, 3

Company data 8×5 converted to the quantitative format 8×7

#	Inco	MSch	NSup	EC	Util	Indu	Reta
1	19.0	43.7	2	0	1	0	0
2	29.4	36.0	3	0	1	0	0
3	23.9	38.0	3	0	0	1	0
4	18.4	27.9	2	1	1	0	0
5	25.7	22.3	3	1	0	1	0
6	12.1	16.9	2	1	0	1	0
7	23.9	30.2	4	1	0	0	1
8	27.2	58.0	5	1	0	0	1

Structure of data at standardization (i), centering

Color/shape corresponds to the product (A,B,C)



This structure has nothing to do with product

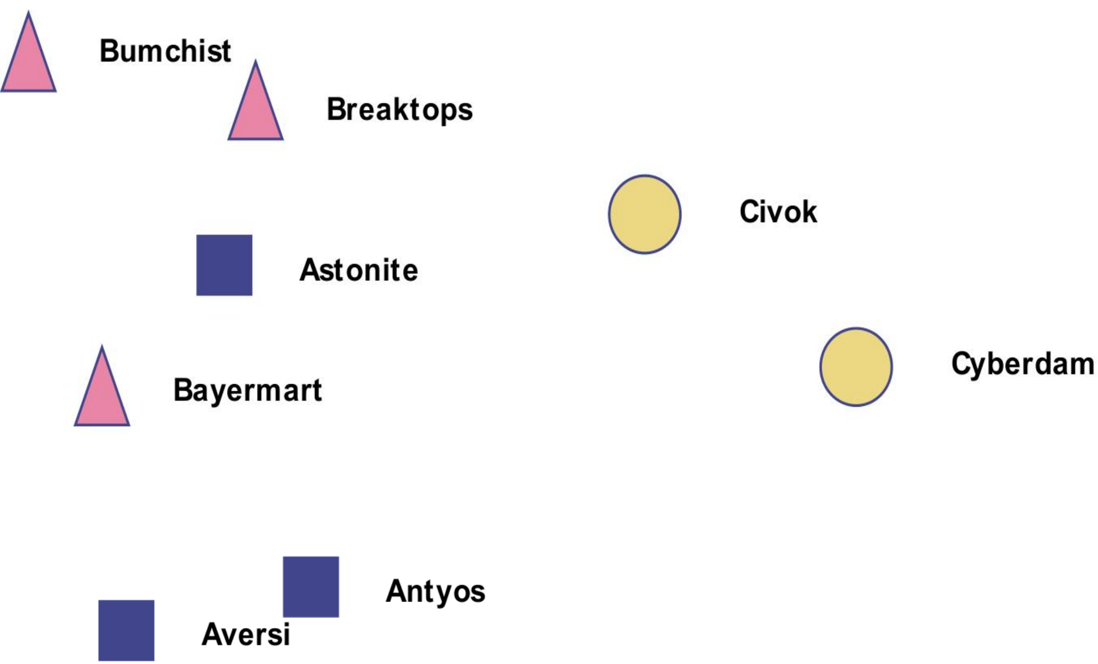
Week 6. Principal Component Analysis: Data standardization issue, 4

Company data 8×5 converted to the quantitative format 8×7

#	Inco	MSch	NSup	EC	Util	Indu	Reta
1	19.0	43.7	2	0	1	0	0
2	29.4	36.0	3	0	1	0	0
3	23.9	38.0	3	0	0	1	0
4	18.4	27.9	2	1	1	0	0
5	25.7	22.3	3	1	0	1	0
6	12.1	16.9	2	1	0	1	0
7	23.9	30.2	4	1	0	0	1
8	27.2	58.0	5	1	0	0	1

Structure of data at standardization (ii), centering and normalizing

Color/shape correspond to the product (A,B,C)



This structure a bit better corresponds to product

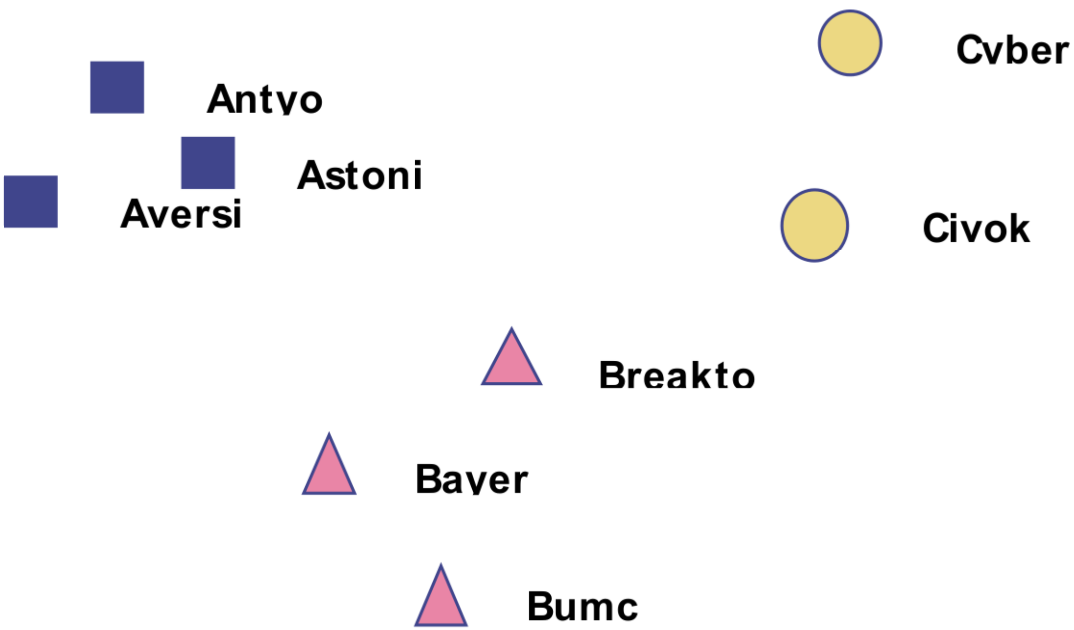
Week 6. Principal Component Analysis: Data standardization issue, 5

Company data 8×5 converted to the quantitative format 8×7

#	Inco	MSch	NSup	EC	Util	Indu	Reta
1	19.0	43.7	2	0	1	0	0
2	29.4	36.0	3	0	1	0	0
3	23.9	38.0	3	0	0	1	0
4	18.4	27.9	2	1	1	0	0
5	25.7	22.3	3	1	0	1	0
6	12.1	16.9	2	1	0	1	0
7	23.9	30.2	4	1	0	0	1
8	27.2	58.0	5	1	0	0	1

Structure of data at standardization (iii), (ii)+ further normalizing Sector features

Color/shape corresponds to the product (A,B,C)



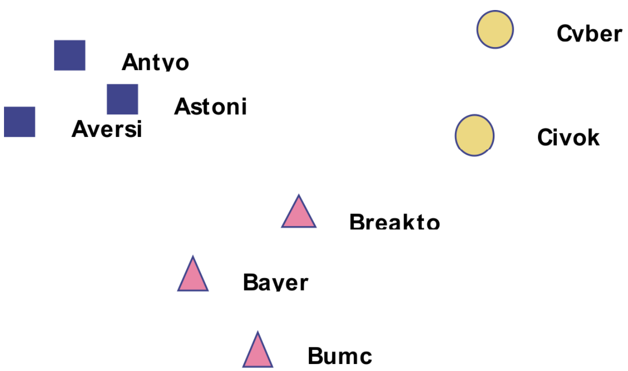
This structure corresponds to product **quite well**

Week 6. Principal Component Analysis: Data standardization issue, 6

Unfortunately, option (iii) sometimes fails too.

The issue of data standardization remains one of the darkest in data analysis.

There have been interesting developments recently – Laplacian transformation and the like.



Yet I think it will remain so unless the human personality structure changes.

Indeed, how can one be sure that their priorities in different aspects of their life are right?

End of Part 1 in week 6

Week 6. Principal Component Analysis 2: Matrix algebra, 1

A matrix (array) N×V is a table of reals with N rows and V columns:

X = $\begin{bmatrix} 6 & 7 \\ 5 & 8 \\ 8 & 9 \end{bmatrix}$**, here N=3, V=2** $\quad$ **X=[x_{iv}]**

X transpose is X turned 90° to the right

X′ = $\begin{bmatrix} 6 & 5 & 8 \\ 7 & 8 & 9 \end{bmatrix}$**, here N=2, V=3** $\quad$ **X′ =[x_{vi}]**

Week 6. Principal Component Analysis 2: Matrix algebra, 2

An ND vector, or ND point, is a matrix N×1

y = $\begin{bmatrix} 6 \\ 5 \\ 8 \end{bmatrix}$**, here N=3**

y's transpose is the y lying down

X′ = $\begin{bmatrix} 6 & 5 & 8 \end{bmatrix}$

Week 6. Principal Component Analysis 2: Matrix algebra, 3

Sum of two vectors, component-wise:

$$\begin{bmatrix} 6 \\ 5 \\ 8 \end{bmatrix} + \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \\ 11 \end{bmatrix}$$

Only vectors of the same dimension may be added to each other.

Week 6. Principal Component Analysis 2: Matrix algebra, 5

Vector multiplied by a number, component-wise

$$a * \begin{bmatrix} 6 \\ 5 \\ 8 \end{bmatrix} = \begin{bmatrix} 6a \\ 5a \\ 8a \end{bmatrix}$$

For example,

$$(-2.5) * \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} = \begin{bmatrix} -2.5 \\ 5.0 \\ -7.5 \end{bmatrix}$$

Week 6. Principal Component Analysis 2: Matrix algebra, 6

$N\times V$ matrix X multiplied by a vector, c (must be $V\times 1$ dimension), Note: a bit unusual formulation:

Xc - is the sum of X's columns weighted by components of c.

For X = $\begin{bmatrix} 6 & 7 \\ 5 & 8 \\ 8 & 9 \end{bmatrix}$ **and c =** $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$

X*c = 2* $\begin{bmatrix} 6 \\ 5 \\ 8 \end{bmatrix}$ **+ 3*** $\begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}$ **=** $\begin{bmatrix} 12 \\ 10 \\ 16 \end{bmatrix}$ **+** $\begin{bmatrix} 21 \\ 24 \\ 27 \end{bmatrix}$ **=** $\begin{bmatrix} 12+21 \\ 10+24 \\ 16+27 \end{bmatrix}$ **=** $\begin{bmatrix} 33 \\ 34 \\ 43 \end{bmatrix}$

Week 6. Principal Component Analysis 2: Matrix algebra, 7

$N\times V$ matrix X multiplied by a matrix, C (must be $V\times M$ dimension),

XC, or $X*C$: the set of columns being products of X and C's columns: if $C=[c_1 \; c_2 \dots c_M]$, then $X*C = C=[X*c_1 \; X*c_2 \dots X*c_M]$

Size: $N\times M$

For $X = \begin{bmatrix} 6 & 7 \\ 5 & 8 \\ 8 & 9 \end{bmatrix}$ and $C = \begin{bmatrix} 2 & 2 \\ 3 & 1 \end{bmatrix}$

$$X*C = [X*\begin{bmatrix} 2 \\ 3 \end{bmatrix} \;\; X*\begin{bmatrix} 2 \\ 1 \end{bmatrix}] = \begin{bmatrix} 33 & 19 \\ 34 & 18 \\ 43 & 25 \end{bmatrix}$$

Week 6. Principal Component Analysis 2: Matrix algebra, 8 : Towards linearity

Consider c = $\begin{bmatrix}2\\3\end{bmatrix}$, 2×1 array,

and 1×50 array a=[0.04 0.08 0.12 0.16 . . . 1.96 2.00]
What is the product c*a?

Answer: c*a=[0.04$\begin{bmatrix}2\\3\end{bmatrix}$ .08$\begin{bmatrix}2\\3\end{bmatrix}$ … 2$\begin{bmatrix}2\\3\end{bmatrix}$]

c*a=$\begin{bmatrix}0.08\\0.12\end{bmatrix}$ $\begin{bmatrix}0.16\\0.24\end{bmatrix}$ … $\begin{bmatrix}4\\6\end{bmatrix}$ = $\begin{bmatrix}0.08 & 0.16 & 0.24 & \ldots 4.00\\0.12 & 0.24 & 0.36 & \ldots 6.00\end{bmatrix}$,

2×50 array.

Week 6. Principal Component Analysis 2: Matrix algebra, 9

Given c = $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and 1×50 array a=[0.04 0.08 0.12 0.16 . . . 1.96 2.00], the product c*a is a 2×50 array of c, variously scaled by factors from 0.04, 0.08, … to 2.

Question: what is the image of this sequence on a Cartesian plane?

For X = $\begin{bmatrix} 6 & 7 \\ 5 & 8 \\ 8 & 9 \end{bmatrix}$, X*c= $\begin{bmatrix} 33 \\ 34 \\ 43 \end{bmatrix}$, what is X*c*a and its image?

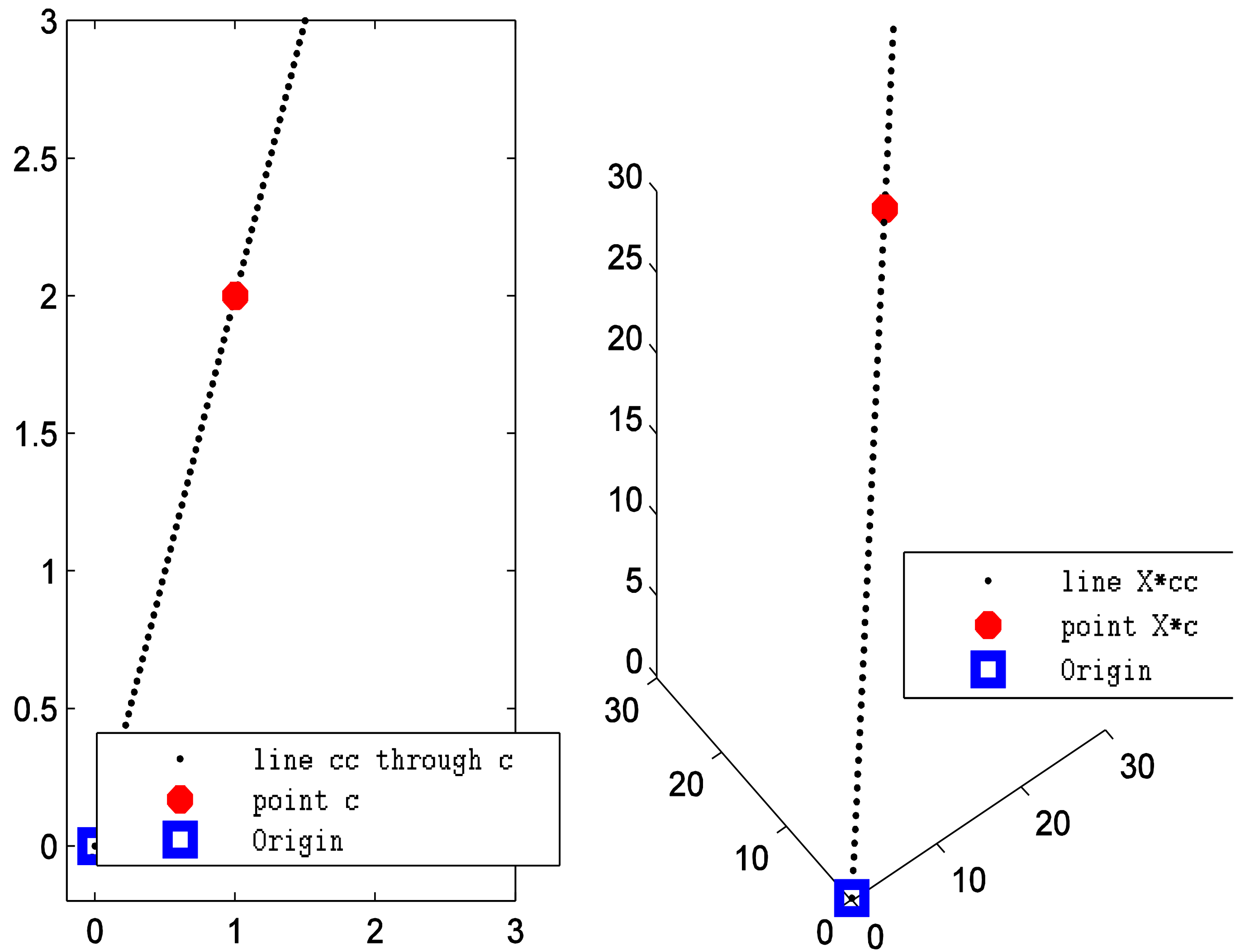
The product X*c*a is a 3×50 array of X*c, variously scaled by factors from 0.04, 0.08, … to 2

Week 6. Principal Component Analysis 2: Matrix algebra, 10: Linearity

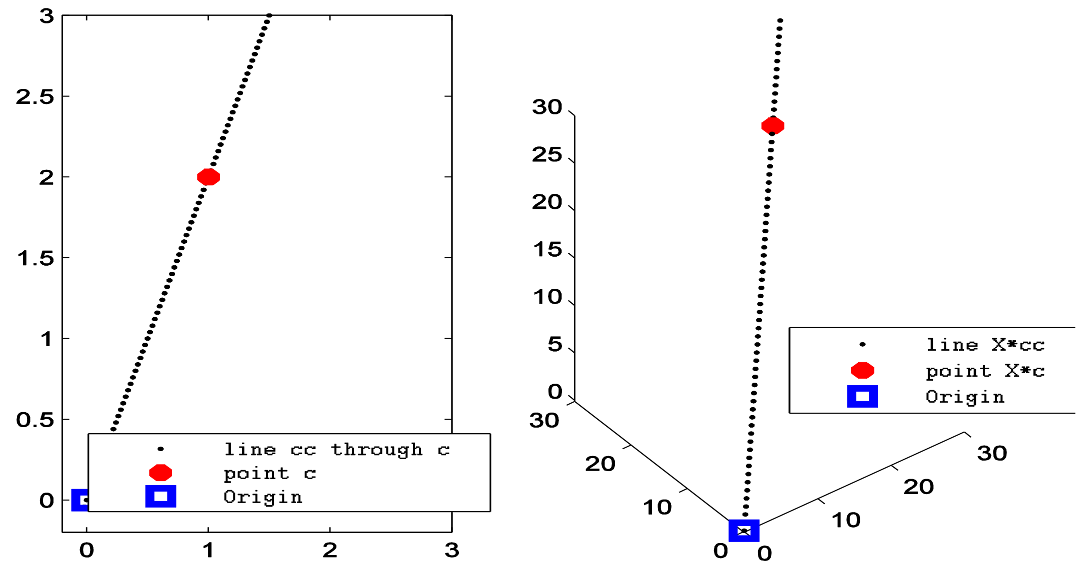
$$X = \begin{bmatrix} 6 & 7 \\ 5 & 8 \\ 8 & 9 \end{bmatrix}$$

and $c = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$:

c*a's image is the 2D line on the left, X*c*a's image is the 3D line on the right.



Week 6. Principal Component Analysis 2: Matrix algebra, 11



True in any dimension: Sequence of scaled values αc, for any vector c, forms a line through c and the origin 0; matrix product X*c scaled by reals also forms a line through X*c and the origin 0. This is why matrix algebra is referred to as linear algebra.

Week 6. Principal Component Analysis2: Dot product, norm, orthogonality, 1

Dot product, or inner product, or scalar product:

<c, d>=c'*d for two m×1 vectors

For

$$c=\begin{bmatrix}1\\2\\3\end{bmatrix}, \qquad d=\begin{bmatrix}-2\\1\\2\end{bmatrix}$$

$$\langle c,d\rangle = [1\ 2\ 3]*\begin{bmatrix}-2\\1\\2\end{bmatrix} = 1*(-2) + 2*1 + 3*2 = -2+2+6=6$$

Dot product of a vector by itself

$$<c, c> = c'*c = \sum_{i=1}^{m} {c_i}^2$$

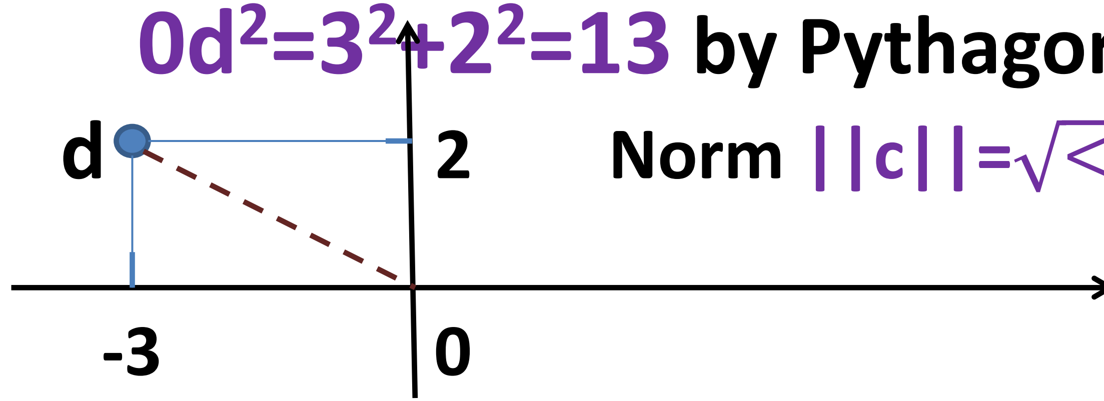
For

$c=\begin{bmatrix}-3\\1\\2\end{bmatrix}$, $<c,c>=(-3)^2+1^2+2^2= 9+1+4=14$

$0d^2=3^2+2^2=13$ by Pythagoras

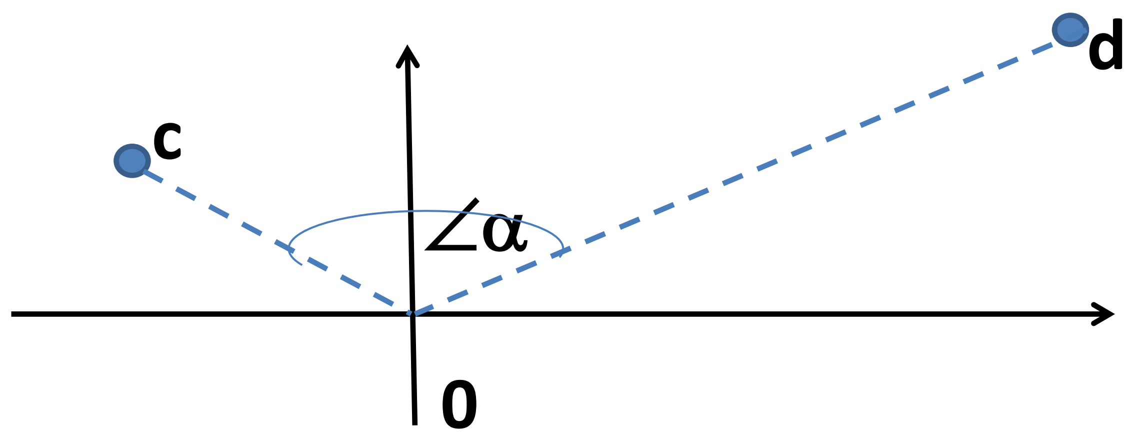
Norm $||c||=\sqrt{< c, c >}$

$d=\begin{bmatrix}-3\\2\end{bmatrix}$



Week 6. Principal Component Analysis 2: Dot product, norm, orthogonality, 3

Geometric meaning of dot product



<c,d> = |c| |d| cos α

c and d are orthogonal if $\angle\alpha$=90° (right angle), then

<c,d>=0

Week 6. Principal Component Analysis 2:

Concepts introduced

Vector, point
Matrix, array
Transpose
Sum of vectors
Vector multiplied by a number
Matrix multiplied by a vector
Matrix multiplication
Linearity
Dot product
Norm
Orthogonality

End of part 2 in Lecture 6

Week 6. Principal Component Analysis 3: Spectrum, 1

Take a square matrix $A=\begin{bmatrix}1 & 1 & 2\\ 1 & 3 & 3\\ 2 & 3 & 1\end{bmatrix}$ **and an** $f=\begin{bmatrix}1\\ -1\\ 1\end{bmatrix}$,

Multiply A*f = $\begin{bmatrix}2\\ 1\\ 0\end{bmatrix}$, **divide A*f./f =** $\begin{bmatrix}2\\ -1\\ 0\end{bmatrix}$ **- differ**

Take c= $\begin{bmatrix}0.3792\\ 0.7171\\ 0.5848\end{bmatrix}$, **multiply A*c=** $\begin{bmatrix}2.2659\\ 4.2849\\ 3.4945\end{bmatrix}$.

Divide

A*c./c= $\begin{bmatrix}5.9754\\ 5.9754\\ 5.9754\end{bmatrix}$ **- same λ=5.9754 :** $A*c= \lambda c$

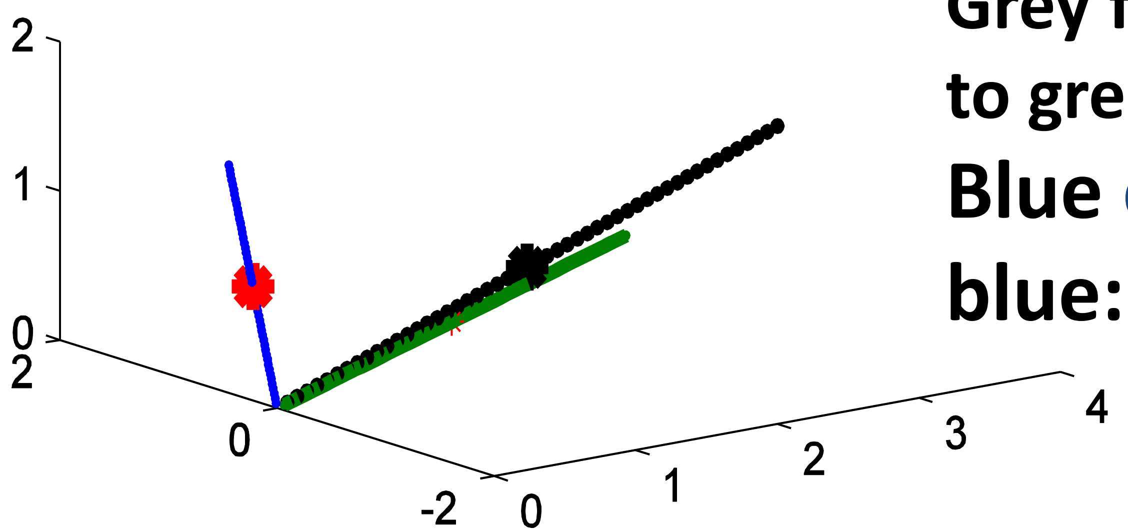
c	eigenvector
λ	eigenvalue

Week 6. Principal Component Analysis 3: Spectrum, 2

For f= $\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$, A*f = $\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$, different direction

For c= $\begin{bmatrix} 0.3792 \\ 0.7171 \\ 0.5848 \end{bmatrix}$, A*c= λc, same direction: eigenvalue and

eigenvector of A

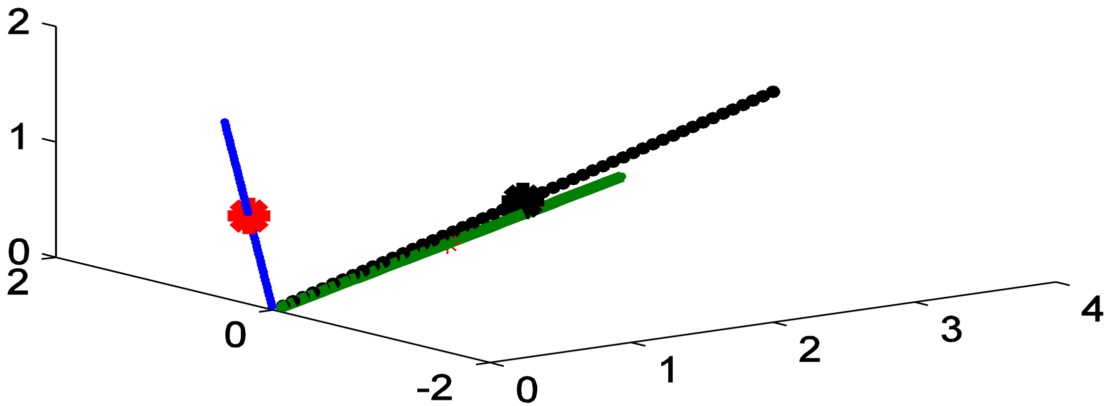


Grey f jumps
to green A*f
Blue c remains on the
blue: eigen axis of A

Week 6. Principal Component Analysis 3: Spectrum, 3

Eigenvalues and eigenvectors: $A^*c = \lambda c$

1. If c is eigenvector for A, then so is αc for any real α; to make c unique, conventionally c is considered normed, |c|=1.
2. The number of different eigenvalues for an m×m A is not greater than m; the number of nonzero eigenvalues is A's rank.
3. If A is symmetric all its eigenvalues are real; and the eigenvectors, mutually orthogonal.



Week 6. Principal Component Analysis 3: Spectrum, 4

Eigen-values/vectors of symmetric matrices: $A*c = \lambda c$

Rank 1 symmetric matrices: A characteristic

Take a vector, say $a=\begin{bmatrix} 1 \\ 2 \\ -4 \end{bmatrix}$, multiply it by its transpose a'=[1 2 -4]:

$$A = \begin{bmatrix} 1 \\ 2 \\ -4 \end{bmatrix} * [1\ 2\ -4] = \begin{bmatrix} 1 & 2 & -4 \\ 2 & 4 & -8 \\ -4 & -8 & 16 \end{bmatrix}$$

In general, for $a=(a_1,a_2,\ldots, a_m)$, matrix A's (i,j)-entry is product

$$a_{ij}=a_i a_j \quad (i,j=1,\ldots,m)$$

Week 6. Principal Component Analysis 3: Spectrum, 5

Eigen-values/vectors of symmetric matrices: $A*c = \lambda c$

Rank 1 symmetric matrices: A characteristic

For A = $\begin{bmatrix} 1 \\ 2 \\ -4 \end{bmatrix}$*[1 2 -4] = $\begin{bmatrix} 1 & 2 & -4 \\ 2 & 4 & -8 \\ -4 & -8 & 16 \end{bmatrix}$, vector *a*= $\begin{bmatrix} 1 \\ 2 \\ -4 \end{bmatrix}$ is its eigenvector:

$A*a$= 1*$\begin{bmatrix} 1 \\ 2 \\ -4 \end{bmatrix}$ +2*$\begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix}$ - 4*$\begin{bmatrix} -4 \\ -8 \\ 16 \end{bmatrix}$ = $\begin{bmatrix} 21 \\ 42 \\ -84 \end{bmatrix}$ =21*$\begin{bmatrix} 1 \\ 2 \\ -4 \end{bmatrix}$ =21*a

at $\lambda = 21 = |a|^2 = \langle a,a \rangle$.

This holds in general too, for matrices $A=(a_i a_j)=aa'$

Week 6. Principal Component Analysis 3: Spectrum, 6

Spectrum of $m \times m$ A: All its eigenvalues, $\lambda_1, \lambda_2, \ldots, \lambda_r$

Respective eigenvectors $c_1, c_2, \ldots, c_r$

Spectral decomposition in three equivalent formats:

(i) vector format $A = \lambda_1 c_1 c_1' + \lambda_2 c_2 c_2' + \cdots \lambda_r c_r c_r'$

Note: $c_1 c_1', \ldots, c_r c_r'$ are rank 1 matrices
Since eigenvectors $c_1, \ldots, c_r$ are mutually orthogonal, these are like cubic building blocks for A

(ii) Entry format $a_{ij} = \lambda_1 c_{i1} c_{j1} + \lambda_2 c_{i2} c_{j2} + \cdots \lambda_r c_{ir} c_{jr}$

Week 6. Principal Component Analysis 3: Spectrum, 7

Spectrum of $m \times m$ A: All its eigenvalues, $\lambda_1, \lambda_2, \ldots, \lambda_r$

Respective eigenvectors $c_1, c_2, \ldots, c_r$

Spectral decomposition in three formats:

(iii) matrix format $A = C\Lambda C'$

where $m \times r$ $C = [c_1 \; c_2 \ldots \; c_r]$ has $c_1, c_2, \ldots, c_r$ as columns, and

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 \ldots & 0 \\ 0 & \lambda_2 \ldots . & 0 \\ & & \ldots . \\ 0 & 0 \ldots & \lambda_r \end{bmatrix}$$ - diagonal matrix of eigenvalues

Footnote: Product of $m \times m$ A and $m \times m$ B is $m \times m$ matrix $C = A*B$ whose columns are products $A*b$ of A by columns b of B.

Week 6. Principal Component Analysis 3: Singular value decomposition, 1 Composition

Take an $N \times V$ data matrix $X=[x_{iv}]$. It transforms any $V \times 1$ vector c into an $N \times 1$ vector z=X*c

X maps VD space into ND space:

At N=3, V=2

X = $\begin{bmatrix} 6 & 7 \\ 5 & 8 \\ 8 & 9 \end{bmatrix}$ and c = $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$, z=X*c = $\begin{bmatrix} 27 \\ 29 \\ 35 \end{bmatrix}$

The transpose, X′, maps inversely, the ND space into the VD space: $\hat{c} = X'z = \begin{bmatrix} 587 \\ 736 \end{bmatrix}$.

Two mappings composed form the product X′X:

$$\hat{c} = X'Xc.$$

Week 6. Principal Component Analysis 3: Singular value decomposition, 2 Composition

Take an $N \times V$ data matrix $X=[x_{iv}]$. It transforms any $V \times 1$ vector c into an $N \times 1$ vector $z=X^*c$

X maps VD space into ND space:

At N=3, V=2: $X = \begin{bmatrix} 6 & 7 \\ 5 & 8 \\ 8 & 9 \end{bmatrix}$ and $c = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$, $z=X^*c = \begin{bmatrix} 27 \\ 29 \\ 35 \end{bmatrix}$

The transpose, X′, maps inversely: $\hat{c} = X'z = \begin{bmatrix} 587 \\ 736 \end{bmatrix}$.

Two mappings form the product X′X:

$$\hat{c} = X'Xc.$$

Is $\hat{c}$ on the same line as c? NO: $\hat{c}./c = \begin{bmatrix} 587 \\ 245.3 \end{bmatrix}$

Can a c be found so that $\hat{c} = X'Xc$ sits on the same line?

Week 6. Principal Component Analysis 3: Singular value decomposition, 3 Composition

$N \times V$ data matrix $X=[x_{iv}]$ transforms VD c into ND $z=X*c$

$V \times N$ transpose $X' =[x_{vi}]$ transforms ND z into VD $\hat{c} = X'*z$

Two mappings form the product $X'X$: $\hat{c} = X'Xc$.

At $X = \begin{bmatrix} 6 & 7 \\ 5 & 8 \\ 8 & 9 \end{bmatrix}$, $X' = \begin{bmatrix} 6 & 5 & 8 \\ 7 & 8 & 9 \end{bmatrix}$ and $X'X = \begin{bmatrix} 125 & 154 \\ 154 & 194 \end{bmatrix}$

Can we find c so that $\hat{c} = X'Xc$ sits on the same line?

Of course!

Week 6. Principal Component Analysis 3: Singular value decomposition, 4 Composition

$N \times V$ data matrix $X=[x_{iv}]$ transforms VD c into ND $z=X*c$

$V \times N$ transpose $X' = [x_{vi}]$ transforms ND z into VD $\hat{c} = X'*z$

At $X = \begin{bmatrix} 6 & 7 \\ 5 & 8 \\ 8 & 9 \end{bmatrix}$, $X' = \begin{bmatrix} 6 & 5 & 8 \\ 7 & 8 & 9 \end{bmatrix}$ and $X'X = \begin{bmatrix} 125 & 154 \\ 154 & 194 \end{bmatrix}$

Can we find c so that $\hat{c} = X'Xc$ sits on the same line, that is, an eigenvector of $X'X$?

Of course! $A = X'X$ is symmetric; it must have two orthogonal eigenvectors so that

$$Ac = \lambda c$$

Week 6. Principal Component Analysis 3: Singular value decomposition, 5 Composition

$N \times V$ data matrix $X=[x_{iv}]$ transforms *VD* c into *ND* $z=X*c$

$V \times N$ transpose $X' =[x_{vi}]$ transforms *ND* z into *VD* $\hat{c} = X'*z$

At $X = \begin{bmatrix} 6 & 7 \\ 5 & 8 \\ 8 & 9 \end{bmatrix}$, $X' = \begin{bmatrix} 6 & 5 & 8 \\ 7 & 8 & 9 \end{bmatrix}$ and $X'X = \begin{bmatrix} 125 & 154 \\ 154 & 194 \end{bmatrix}$

$A = X'X$ is symmetric; it must have two orthogonal eigenvectors satisfying

$$Ac=\lambda c$$

Moreover, $X'X$ has no negative eigenvalues since it is “quadratic” (positive semi definite)

Week 6. Principal Component Analysis 3: Singular value decomposition, 6 Composition

$N \times V$ data matrix $X=[x_{iv}]$ transforms VD c into ND $z=X*c$

$V \times N$ transpose $X' =[x_{vi}]$ transforms ND z into VD $\hat{c} = X'*z$

At $X = \begin{bmatrix} 6 & 7 \\ 5 & 8 \\ 8 & 9 \end{bmatrix}$, $X' = \begin{bmatrix} 6 & 5 & 8 \\ 7 & 8 & 9 \end{bmatrix}$ and $A=X'X= \begin{bmatrix} 125 & 154 \\ 154 & 194 \end{bmatrix}$

Matlab:

```
>> [C, La]=eig(A)
```

$$C = \begin{bmatrix} -0.7806 & 0.6251 \\ 0.6251 & 0.7806 \end{bmatrix}$$

Columns of C are eigenvectors

$$La= \begin{bmatrix} 1.68 & 0 \\ 0 & 317.32 \end{bmatrix}$$

Positive eigenvalues

Week 6. Principal Component Analysis 3: Singular value decomposition, 7 Composition

$N\times V$ data matrix $X=[x_{iv}]$ transforms VD c into ND $z=X*c$

$V\times N$ transpose $X'=[x_{vi}]$ transforms ND z into VD $\hat{c} = X'*z$

At $X = \begin{bmatrix} 6 & 7 \\ 5 & 8 \\ 8 & 9 \end{bmatrix}$, $X' = \begin{bmatrix} 6 & 5 & 8 \\ 7 & 8 & 9 \end{bmatrix}$ and $A=X'*X= \begin{bmatrix} 125 & 154 \\ 154 & 194 \end{bmatrix}$

Found $c = \begin{bmatrix} 0.6251 \\ 0.7806 \end{bmatrix}$ and $\lambda = 317.32$

so that c and $A*c$ sit on the same line: $A*c = \lambda c$

$$\begin{bmatrix} 125 & 154 \\ 154 & 194 \end{bmatrix} * \begin{bmatrix} 0.6251 \\ 0.7806 \end{bmatrix} = 317.32 * \begin{bmatrix} 0.6251 \\ 0.7806 \end{bmatrix}$$

Week 6. Principal Component Analysis 3: Singular value decomposition, 1 Singular triplet

$N \times V$ data matrix $X=[x_{iv}]$ transforms VD c into ND z=X*c

$V \times N$ transpose $X'=[x_{vi}]$ transforms ND z into VD $\hat{c} = X'$*z

$$A=X'*X$$

There is $\lambda > 0$ $A*c = \lambda c$: eigen-couple (λ, c)

Let us present this in terms of X:

$$X'*(X*c) = \lambda c = \mu^2 c \quad \text{where} \quad \mu = \sqrt{\lambda}$$

$$X'*(X*c/\mu) = \mu c$$

$$X'*z = \mu c, \quad z = X*c/\mu$$

A triplet (μ, c, z) is singular for X:

$$\begin{cases} X*c = \mu z \\ X'*z = \mu c \end{cases} \quad (*)$$

Week 6. Principal Component Analysis 3: Singular value decomposition, 2 Singular triplet

$N \times V$ data matrix $X=[x_{iv}]$ transforms *VD* c into *ND* $z=X*c$

$V \times N$ transpose $X' = [x_{vi}]$ transforms *ND* z into *VD* $\hat{c} = X'*z$

A triplet (μ, c, z) is singular for X, that is,

$$\begin{cases} X*c = \mu z \\ X'*z = \mu c \end{cases} \qquad (*)$$

if and only if c is eigenvector of $A=X'*X$ corresponding to eigenvalue $\lambda = \mu^2$,

and $z = X*c/\mu$

Week 6. Principal Component Analysis 3: Singular value decomposition, 3 Singular triplets

$N \times V$ data matrix $X=[x_{iv}]$ transforms VD c into ND z=X*c

$V \times N$ transpose X' =$[x_{vi}]$ transforms ND z into VD $\hat{c} = X'$*z

(μ, c, z) X-singular $\Leftrightarrow$ c is eigenvector of $A=X'*X$ with eigenvalue $\lambda = \mu^2$

Inherited from the spectral analysis:

1. The number of different singular triplets is equal to the rank of X (=rank of $X'*X$);

2. Different singular c are mutually orthogonal as well as different singular z.

Week 6. Principal Component Analysis 3: Singular value decomposition, 1 SVD

Data X: The singular values sorted $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_r$

Respective singular VD vectors $c_1, \quad c_2, \ldots, \quad c_r$

Respective singular ND vectors $z_1, \quad z_2, \ldots, \quad z_r$

(all normed)

SVD in three equivalent formats:

(i) Vector format $$X = \mu_1 z_1 c_1' + \mu_2 z_2 c_2' + \cdots \mu_r z_r c_r'$$

Note: $z_1 c_1', \ldots, z_r c_r'$ are rank 1 matrices

(ii) Entry format $$x_{iv} = \mu_1 z_{i1} c_{v1} + \mu_2 z_{i2} c_{v2} + \cdots \mu_r z_{ir} c_{vr}$$

Week 6. Principal Component Analysis 3: Singular value decomposition, 2 SVD

Data X: The singular values sorted $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_r$

Respective singular VD vectors $c_1, \quad c_2, \ldots, \quad c_r$ (normed)

Respective singular ND vectors $z_1, \quad z_2, \ldots, \quad z_r$ (normed)

SVD in three equivalent formats:

(iii) Matrix format $A=ZMC'$

$V \times r$ $C=[c_1 \;\; c_2 \ldots \;\; c_r]$ has $c_1, \;\; c_2, \ldots, \;\; c_r$ as columns,

$N \times r$ $Z=[z_1 \;\; z_2 \ldots \;\; z_r]$ has $z_1, \;\; z_2, \ldots, \;\; z_r$ as columns,

and

$$M=\begin{bmatrix} \mu_1 & 0 \ldots & 0 \\ 0 & \mu_2 \ldots & 0 \\ & & \ldots \\ 0 & 0 \ldots & \mu_r \end{bmatrix}$$ - diagonal matrix of singular values

Week 6. Principal Component Analysis 3: Singular value decomposition: Approximation 1

Data X**: The singular values sorted** $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_r$

SVD: $X = \mu_1 z_1 c_1' + \mu_2 z_2 c_2' + \cdots \mu_r z_r c_r'$

Rank is a mathematical explication of the space dimension.

Problem: Given X=[x_{iv}], find a matrix Xp=[xp_{iv}] of rank $p < r$ minimizing the sum-of-squares difference

$$\|X - Xp\|^2 = \sum_{i,v} (x_{iv} - xp_{iv})^2$$

Solution: The first p singular triplets,

$$Xp = \mu_1 z_1 c_1' + \mu_2 z_2 c_2' + \cdots \mu_p z_p c_p'$$

Week 6. Principal Component Analysis 3: Singular value decomposition: Approximation 2

Data X: The singular values sorted $\mu_1 \geq \mu_2 \geq \ldots \geq \mu_r$

SVD: $X = \mu_1 z_1 c_1' + \mu_2 z_2 c_2' + \cdots \mu_r z_r c_r'$

Matrix

$$Xp = \mu_1 z_1 c_1' + \mu_2 z_2 c_2' + \cdots \mu_p z_p c_p', \ p < r$$

minimizes the sum-of-squares difference

$\|X - Xp\|^2 = \sum_{i,v}(x_{iv} - xp_{iv})^2$ over all matrices of rank p

Data scatter decomposition

$$\|X\|^2 = \mu_1^2 + \mu_2^2 + \ldots + \mu_p^2 + \|X - Xp\|^2$$

where $\|X\|^2 = \sum_{i,v} x_{iv}^2$, the data scatter;

μ_k^2 - the contribution of k-th singular triplet

End of Part 3 in Week 6

Week 6. Principal Component Analysis 4: Hidden Factor Model, 1

#	SEn	OOP	CI	Average
1	41	66	90	65.7
2	57	56	60	57.7
3	61	72	79	70.7
4	69	73	72	71.3
5	63	52	88	67.7
6	62	83	80	75.0

F. Galton: Set of students, marks over Software engineering, Object-Oriented Programming, Computational Intelligence

Can table help in deriving hidden talent scores?

Why would one want anything better than the average? (Sorry, no answer till next week.)

Mark(Stud., Subj.)=Talent_Score(Student)∗Loading(Subject)

Type of Model: Observed=F(Hidden)+Error

Week 6. Principal Component Analysis 4: Hidden Factor Model, 2

#	SEn	OOP	CI	Average
1	41	66	90	65.7
2	57	56	60	57.7
3	61	72	79	70.7
4	69	73	72	71.3
5	63	52	88	67.7
6	62	83	80	75.0

Given marks x_{iv}, model $\boldsymbol{x_{iv} = z_i * c_v + e_{iv}}$

$$L = \sum_{i=1}^{N} \sum_{v=1}^{V} (x_{iv} - z_i c_v)^2 \Rightarrow \min_{z,\, c}$$

z_i- Hidden factor (Talent) score

c_v- Subject loading (the easier the subject, the higher the load)

Week 6. Principal Component Analysis: Hidden Factor Model, 3

#	SEn	OOP	CI	Average
1	41	66	90	65.7
2	57	56	60	57.7
3	61	72	79	70.7
4	69	73	72	71.3
5	63	52	88	67.7
6	62	83	80	75.0

$$\sum_{i,v} x_{iv}{}^2 = \mu_1{}^2 + L \quad \textbf{(2)}$$

Data scatter decomposed in the contribution and criterion

$$L = \sum_{i=1}^{N} \sum_{v=1}^{V} (x_{iv} - z_i c_v)^2 \Rightarrow \min_{z,\, c}$$

Solution up to the product, First singular triplet

$$\boldsymbol{x_{iv} \approx \mu_1 z_{i1} c_{v1}} \qquad \textbf{(1)}$$

where μ_1 is the maximum singular value of X

Week 6. Principal Component Analysis: Hidden Factor Model, 4

#	SEn	OOP	CI	Average
1	41	66	90	65.7
2	57	56	60	57.7
3	61	72	79	70.7
4	69	73	72	71.3
5	63	52	88	67.7
6	62	83	80	75.0

$$\sum_{i,v} {x_{iv}}^2 = {\mu_1}^2 + L \quad \textbf{(2)}$$

Availability of the contribution!

At model $\boldsymbol{x_{iv} = z_i c_v}$, the best least squares approximation is the First singular triplet

$$\boldsymbol{x_{iv} \approx \mu_1 z_{i1} c_{v1}} \qquad \textbf{(1)}$$

Therefore, least squares solution to the model, the **principal component score and loading**:

$$\boldsymbol{z_i = \sqrt{\mu_1} z_{i1}}$$

$$\boldsymbol{c_v = \sqrt{\mu_1} c_{v1}}$$

Week 6. Principal Component Analysis: Hidden Factor Model, 5

* =

#	SEn	OOP	CI	Average
1	41	66	90	65.7
2	57	56	60	57.7
3	61	72	79	70.7
4	69	73	72	71.3
5	63	52	88	67.7
6	62	83	80	75.0

How from here to move to a 100 grade factor scale? Will be seen next week.

$$
\begin{bmatrix} 41 & 66 & 90 \\ 57 & 56 & 60 \\ 61 & 72 & 79 \\ 69 & 73 & 72 \\ 63 & 52 & 88 \\ 62 & 83 & 80 \end{bmatrix}
=
\begin{bmatrix} 6.85 \\ 5.83 \\ 7.21 \\ 7.20 \\ 6.95 \\ 7.64 \end{bmatrix}
\begin{bmatrix} 8.45 & 9.67 & 11.25 \end{bmatrix}
+
\begin{bmatrix} -16.88 & -0.29 & 12.93 \\ 7.78 & -0.37 & -5.53 \\ 0.12 & 2.28 & -2.05 \\ 8.17 & 3.33 & -9.00 \\ 4.31 & -15.22 & 9.85 \\ -2.53 & 9.09 & -5.92 \end{bmatrix}
$$

Factor scores

Loadings

Residuals

Week 6. Principal Component Analysis: Hidden Factor Model, 6

* =

#	SEn	OOP	CI	Average
1	41	66	90	65.7
2	57	56	60	57.7
3	61	72	79	70.7
4	69	73	72	71.3
5	63	52	88	67.7
6	62	83	80	75.0

MatLab code for

$$X=\begin{bmatrix} 41 & 66 & 90 \\ 57 & 56 & 60 \\ 61 & 72 & 79 \\ 69 & 73 & 72 \\ 63 & 52 & 88 \\ 62 & 83 & 80 \end{bmatrix}$$

```
>> [Z,Mu,C]=svd(X); % unfortunately, MatLab's output is excessive
>> Z=Z(:,1:3)          % three singular 6D scoring vectors
>> Mu=Mu(1:3,:)        % three singular values on diagonal
```

Week 6. Principal Component Analysis: Hidden Factor Model, 7

$$X=\begin{bmatrix} 41^* & 66 & 90 \\ 57 & 56 & 60 \\ 61 & 72 & 79 \\ 69 & 73 & 72 \\ 63 & 52 & 88 \\ 62 & 83 & 80 \end{bmatrix}$$

```
>> [Z,Mu,C]=svd(X); % unfortunately, MatLab output is excessive
>> Z=Z(:,1:3)          % three singular 6D scoring vectors
>> Mu=Mu(1:3,:)        % three singular values on diagonal
```

	First	Second	Third
	-0.4015	-0.7139	-0.4033
	-0.3414	0.3114	0.2075
Z=	-0.4222	0.0859	-0.0948
	-0.4219	0.4583	0.0582
	-0.4071	-0.3594	0.7596
	-0.4476	0.2166	-0.4528
μ=	291.39	27.37	20.84
	-0.4948	0.5895	0.6385
C=	-0.5667	0.3382	-0.7513
	-0.6588	-0.7336	0.1667

Pitfall: $\mathbf{z}_1$ and $\mathbf{c}_1$ are negative

Way out: Take $-\mathbf{z}_1$ and $-\mathbf{c}_1$ instead

Indeed:

$$\begin{cases} X*c = \mu z \\ X'*z = \mu c \end{cases}$$

If and only if

$$\begin{cases} X*(-c) = \mu(-z) \\ -X'*(-z) = -\mu(-c) \end{cases}$$

Week 6. Principal Component Analysis: Hidden Factor Model, 8

$$X=\begin{bmatrix} 41^* & 66 & 90 \\ 57 & 56 & 60 \\ 61 & 72 & 79 \\ 69 & 73 & 72 \\ 63 & 52 & 88 \\ 62 & 83 & 80 \end{bmatrix}$$

>> [Z,Mu,C]=svd(X); % unfortunately, MatLab output is excessive
>> Z=Z(:,1:3) % three singular 6D scoring vectors
>> Mu=Mu(1:3,:) % three singular values on diagonal
>>ds=sum(sum(X.*X)) % data scatter

	First	Second	Third
	0.4015	**-0.7139**	**-0.4033**
	0.3414	**0.3114**	**0.2075**
Z=	**0.4222**	**0.0859**	**-0.0948**
	0.4219	**0.4583**	**0.0582**
	0.4071	**-0.3594**	**0.7596**
	0.4476	**0.2166**	**-0.4528**
μ=	**291.39**	**27.37**	**20.84**
	0.4948	**0.5895**	**0.6385**
C=	**0.5667**	**0.3382**	**-0.7513**
	0.6588	**-0.7336**	**0.1667**

Changed signs in z_1 and **c_1**

Contribution of first component
μ_1^2=291.39^2 = 84909 to
Data scatter ds=86092
is

$$\mu_1^2/\text{ds}= 0.9863=98.6\%$$

Week 6. Principal Component Analysis: Hidden Factor Model, 9

Y= * =

$$\begin{bmatrix} -17.83 & -1.00 & 11.83 \\ -1.83 & -11.00 & -18.17 \\ 2.17 & 5.00 & 0.83 \\ 10.17 & 6.00 & -6.17 \\ 4.17 & -15.00 & 9.83 \\ 3.17 & 16.00 & 1.83 \end{bmatrix}$$

What happens if data is preprocessed by centering, subtracting the column means from the columns?

```
>>Y=X-repmat(mean(X),6,1); % centering
>> [Z,Mu,C]=svd(Y);
```

	First	Second	Third
	-0.7086	0.1783	0.4534
	0.2836	-0.6934	0.4706
	0.0935	0.1841	-0.0486
Z=	0.4629	0.0931	-0.1916
	-0.3705	-0.3374	-0.7293
	0.2391	0.5753	0.0455
μ=	27.37	26.13	17.26
	0.5933	-0.0056	-0.8049
C=	0.3734	0.8878	0.2690
	-0.7131	0.4601	-0.5289

(1) The principal component drastically changes

(2) Contribution of first component changes too

μ_1^2=27.37^2 = 749.1 to

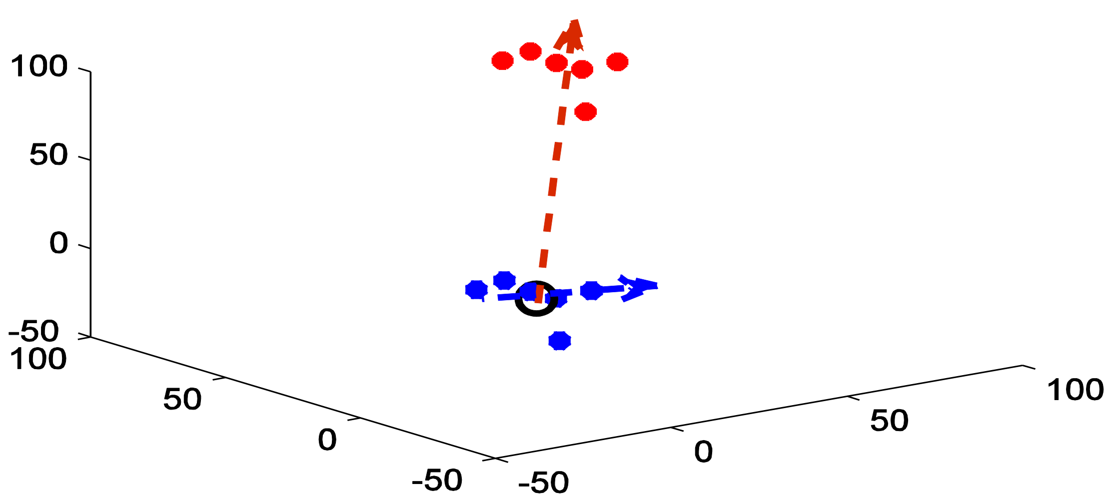
Data scatter ds=1729.7 is μ_1^2/ds=0.433, down to 43.3% from 98.6% previously

Week 6. Principal Component Analysis: Hidden Factor Model, 10

* =

Why such a drastic change? after the data X is preprocessed into Y by centering, that is, subtracting the column means from the columns

Contribution down to 43.3% from 98.6%, etc.



Red dots – raw data X

Blue – data Y centered

Circle – space origin

Red arrow – PC for X

Blue arrow – PC for Y

Why? Because all PCs must go through the origin.

Week 6. Principal Component Analysis Hidden Factor Model, 11

* =

PCA methodology: Summary

- Principal component is a rescaled singular triplet
- Principal components optimally approximate the data according to the least squares criterion
- Principal components are orthogonal to each other
- Principal component's contribution is proportional to its squared singular value
- Principal components change at any data transformation, centering included
- Principal components of X are highly related to eigenvalues and eigenvectors of A=X′X

End of Part 4 in Week 6

Week 6. Principal Component Analysis 5: Conventional approach, 1

Conventional definition:

Given a data matrix *X*, first principal component (PC):

a weighted combination *z* of *X* features after centering, that is, $z=X*c$,

that takes into account as great part of the sum of all feature variances as possible with respect to all normed *c*

Second PC is defined similarly except for a condition that it must be orthogonal to the First PC; Third PC must be orthogonal to both first and second PCs, etc.

Week 6. Principal Component Analysis 5: Conventional approach, 2

Computing the First Principal Component:

1. Given a N×V data matrix *X*, compute its centered version Y and the V×V feature covariance matrix B:
2. Find the first eigenvalue λ_1 and corresponding normed eigenvector c_1 so that $Bc_1 = \lambda_1 c_1$;
3. Compute the principal component

$$z = \frac{Yc_1}{\sqrt{N\lambda_1}}$$

The 2^{nd} PC is computed in the same way based on a residual covariance matrix $B = B - \lambda_1 c_1 c_1'$, etc.

Week 6. Principal Component Analysis 5: Conventional approach, 3

Covariance matrix:

1. Given a N×V data matrix ***X***, compute its centered version Y and the V×V feature covariance matrix B:
   a. Center matrix X by finding, for each feature, its mean and subtracting it from all the feature values, Y=X-m(X)
   b. Compute square matrix A=Y′*Y and divide it by N or N-1 (do the latter if you think that the result is going to be used as an estimate of the covariance matrix of a multivariate density function, I rather divide by N): B= Y′*Y/N.

(*v*,*w*) entry in B: $b_{vw} = \frac{1}{N}\sum_{i=1}^{N}(x_{iv} - \overline{x_v})(x_{iw} - \overline{x_w})$

Week 6. Principal Component Analysis 5: Conventional approach, 4

* =

Covariance matrix:

Given a N×V data matrix *X*, its V×V feature covariance matrix $B=[b_{vw}]$:

$$b_{vw} = \frac{1}{N}\sum_{i=1}^{N}(x_{iv} - \overline{x_v})(x_{iw} - \overline{x_w}),$$

$\overline{x_v}, \overline{x_w}$ - means

Correlation matrix:

If features in a N×V data matrix *X*, have been normalized by their standard deviations, then the covariances b_{vw} are correlation coefficients

$$b_{vw} = \frac{1}{\sigma_v \sigma_w}\sum_{i=1}^{N}(x_{iv} - \overline{x_v})(x_{iw} - \overline{x_w}),$$

σ_v, σ_w - standard deviations

Week 6. Principal Component Analysis 5: Conventional approach, 5 Example

1. Given X and its centered version Y,
compute the covariance matrix **B= Y′∗Y/N.**

$$X=\begin{bmatrix} 41 & 66 & 90 \\ 57 & 56 & 60 \\ 61 & 72 & 79 \\ 69 & 73 & 72 \\ 63 & 52 & 88 \\ 62 & 83 & 80 \end{bmatrix} \quad Y=\begin{bmatrix} -17.83 & -1.00 & 11.83 \\ -1.83 & -11.00 & -18.17 \\ 2.17 & 5.00 & 0.83 \\ 10.17 & 6.00 & -6.17 \\ 4.17 & -15.00 & 9.83 \\ 3.17 & 16.00 & 1.83 \end{bmatrix} \quad B=\begin{bmatrix} 76.14 & 16.33 & -31.97 \\ 16.33 & 110.67 & 6.17 \\ -31.97 & 6.17 & 101.47 \end{bmatrix}$$

2. Compute first eigenvalue and eigenvector of **B= Y′∗Y/N:**

>>[C,La]=eig(B);% eigenvalues in the descending order

>>l1=La(3,3) % λ1=124.85

>>c1=C(:,3);% $c1=\begin{bmatrix} -0.59 \\ -0.37 \\ 0.71 \end{bmatrix}$

Week 6. Principal Component Analysis 5: Conventional approach, 6 Example

* =

Y=

$$\begin{bmatrix} -17.83 & -1.00 & 11.83 \\ -1.83 & -11.00 & -18.17 \\ 2.17 & 5.00 & 0.83 \\ 10.17 & 6.00 & -6.17 \\ 4.17 & -15.00 & 9.83 \\ 3.17 & 16.00 & 1.83 \end{bmatrix}$$

λ1=124.85

c1= $\begin{bmatrix} \mathbf{-0.59} \\ \mathbf{-0.37} \\ \mathbf{0.71} \end{bmatrix}$

3. Given **centered data Y, eigenvalue λ1 and eigenvector c1,** compute the Principal component scoring vector

$$\boldsymbol{z} = \frac{\boldsymbol{Yc_1}}{\sqrt{\boldsymbol{6*\lambda_1}}}$$

Z= $\begin{bmatrix} 0.71 \\ -0.28 \\ -0.09 \\ -0.46 \\ 0.37 \\ -0.24 \end{bmatrix}$

>>z=Y*c1/sqrt(6*λ1);

Week 6. Principal Component Analysis 5: Conventional approach, 7 Example

Y=

$$\begin{bmatrix} -17.83 & -1.00 & 11.83 \\ -1.83 & -11.00 & -18.17 \\ 2.17 & 5.00 & 0.83 \\ 10.17 & 6.00 & -6.17 \\ 4.17 & -15.00 & 9.83 \\ 3.17 & 16.00 & 1.83 \end{bmatrix}$$

Solution:

λ1=124.85

c1= $\begin{bmatrix} -0.59 \\ -0.37 \\ 0.71 \end{bmatrix}$

Z= $\begin{bmatrix} 0.71 \\ -0.28 \\ -0.09 \\ -0.46 \\ 0.37 \\ -0.24 \end{bmatrix}$

Compare to the solution found on slide 57 using the model-based approach:

μ= 27.37

C= $\begin{matrix} \mathbf{0.5933} \\ \mathbf{0.3734} \\ \mathbf{-0.7131} \end{matrix}$

Z= $\begin{bmatrix} -0.7086 \\ 0.2836 \\ 0.0935 \\ 0.4629 \\ -0.3705 \\ 0.2391 \end{bmatrix}$

Well: (c1,z) coincide with (c,z) up to the rounding error and multiplying by -1; 27.37= $\sqrt{6 * 124.85}$. **The same!!! Why?**

Week 6. Principal Component Analysis 5: Conventional approach, 8

* =

Why leads the conventional PCA approach to the same scoring and loading vectors as the model-based PCA?
The former operates over the covariance matrix never used by the latter.

Because the covariance matrix coincides, up to a constant factor, with matrix A=X′∗X, provided that X is centered. **Matrix A is in the core of Singular triplets.**

Working with eigenvectors of A is equivalent to working with singular vectors of X, as proven in Part 3, slide 42.

Week 6. Principal Component Analysis 5: Conventional approach, 9

*   =

The conventional PCA method gives the same results that the model-based PCA applied to the centered version of the data matrix.

What is the difference then?

Week 6. Principal Component Analysis 5:
Difference between model-based and Conventional approaches

- Model-based PCA models data, whereas conventional PCA is purely heuristic.
- Model-based PCA derives that a PCA scoring vector is a weighted sum of the features, whereas conventional PCA presumes that.
- Model-based PCA applies to any data-preprocessing option, whereas conventional PCA needs features centered
- Model-based PCA gives contributions to the data scatter, whereas conventional PCA does not
- Model-based PCA involves approximation of the data by a lower rank space in which a further search for a "base of simple structure" is possible, whereas the conventional PCA can use only bases consisting of eigenvectors

Week 6. Principal Component Analysis: Method and Model

Summary

- Theoretic introduction: Summarization versus Correlation: **Summarization is similar to correlation if all features are considered target; the data standardization issue is important yet to be explored.**
- Matrix operations: **Be cautious, matrices are not numbers!**
- Matrix spectrum, singular value decomposition, approximation: **Considering a square matrix as a mapping, eigenvectors represent axes that are mapped onto themselves. For symmetric matrices these axes are mutually orthogonal. A rectangular matrix can be treated similarly if both direct and inverse mapping are considered. Luckily this amounts to multiplication of matrices and using one more, semi positive definite, property.**

Week 6. Principal Component Analysis: Method and Model

Summary

- Hidden factor model. Its solution. Principal components, loadings, contributions. **This all comes from the SVD theory because of somewhat over simplistic character of the model, just a product of row-related item and a column-related item.**
- Conventional PCA criterion and method. Relation between the model-based and conventional approaches. Covariance and correlation matrix. **Amazingly, the covariance and correlation matrices are closely related to a matrix of direct-inverse mapping, which provides for getting the same results. Yet the model-based criterion has a number of good properties that are going to be used next week.**