STEVENS
INSTITUTE of TECHNOLOGY
School of Business

Amir H Gandomi; PhD
Assistant Professor
Stevens Institute of Technology
a.h.gandomi@stevens.edu

Multivariate Data Analysis – BIA 652

Class 9 – Principal Component Analysis & Singular Value Decomposition

Overview of Class 9

- Overview of Classification/Clustering.
- Principal Component Analysis – Chapter 14
- Singular Value Decomposition
- Review of some Matrix Analysis
- HW : PCA analysis of a dataset
- Important Dates:
  - Due date to submit Poster draft: **April 10, 2019**
  - Late term exam (in class for BIA-652-A): **April 15, 2019**
  - Submitting the final version of your poster: **April 20, 2019**
  - In class and oral presentation: **April 22 and 29, 2019**
  - Poster Event (Corporate Networking Event): **April 30, 2019**
  - Deadline to submit your complete project report: **May 6, 2019**

Where we are:

- If there is a continuous outcome variable:
  - Perform Multiple Linear Regression
- If there is an outcome variable:
  - Perform a Classification, e.g. logistic regression or discriminant analysis or kNN or Gaussian Naïve Bayes
- To group observations:
  - Perform Cluster Analysis: e.g. hierarchical, k-means, density
- (Now) To restructure a group of variables to reduce the number of dimensions (and for many other applications):
  - Perform PCA, SVD or Factor Analysis

A List of Classification Algorithms
(From Wikipedia, but not bad)

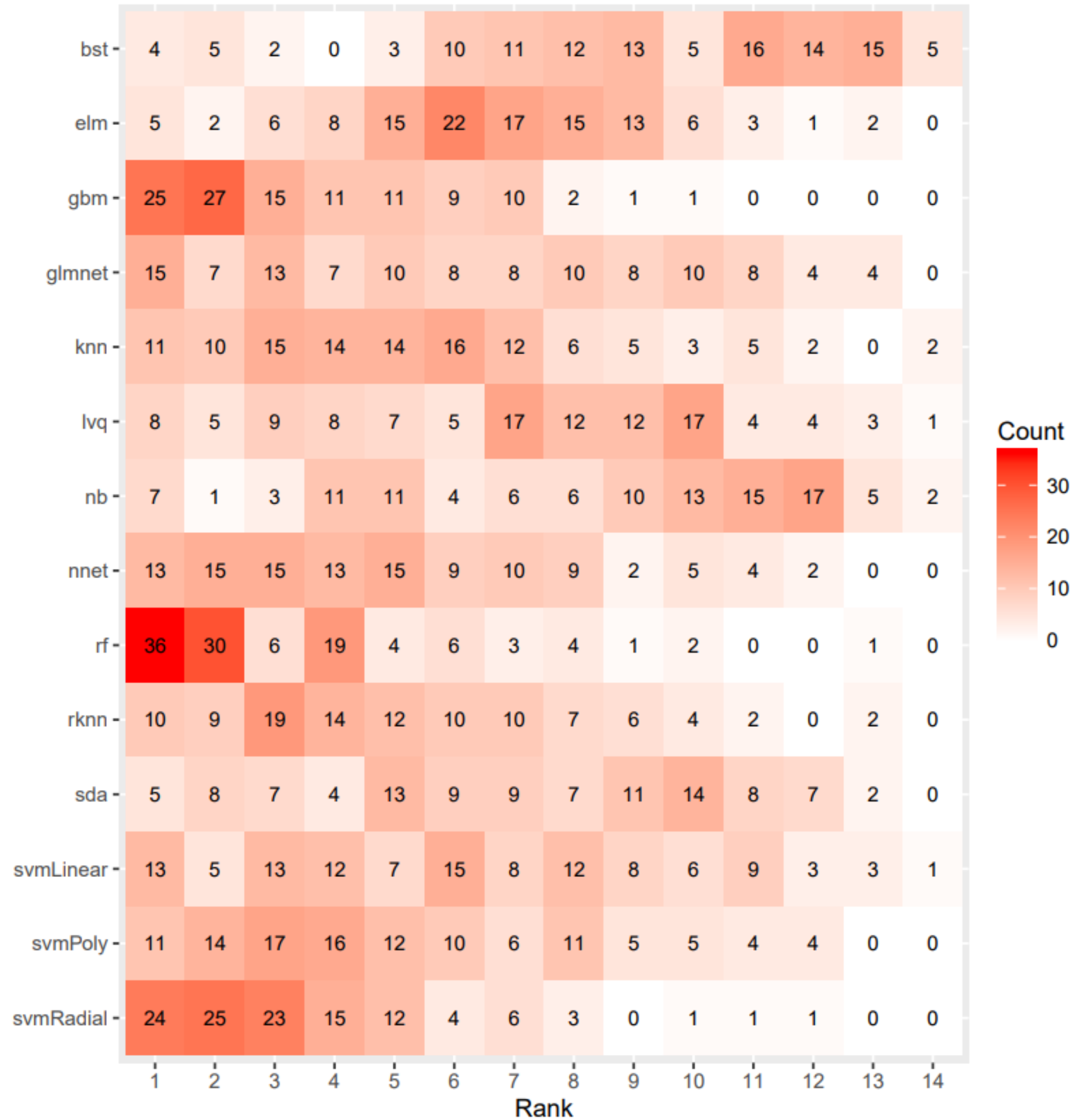
- Linear classifiers
  - Fisher's linear discriminant
  - Logistic regression
  - Naive Bayes classifier
  - Perceptron
- Support vector machines
  - Least squares support vector machines
- Quadratic classifiers
- Kernel estimation
  - k-nearest neighbor
- Boosting (meta-algorithm)
- Decision trees
  - Random forests
- Neural networks

Some Additional Classification Sources (A bit ML oriented, and by no means complete)

1. Wainer, J. (2016). Comparison of 14 different families of classification algorithms on 115 binary datasets. *arXiv preprint arXiv:1606.00930*. https://arxiv.org/pdf/1606.00930.pdf
2. Lim, T. S., Loh, W. Y., & Shih, Y. S. (2000). A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine learning*, *40*(3), 203-228. https://doi.org/10.1023/A:100760822
3. SURVEY PAPER Top 10 algorithms in data mining http://www.cs.umd.edu/~samir/498/10Algorithms-08.pdf
4. Essentials of Machine Learing Algorithms (with Python and R code) http://www.analyticsvidhya.com/blog/2015/08/common-machine-learning-algorithms/
5. A Simple Guide to Logistic Regression in R. http://www.analyticsvidhya.com/blog/2015/11/beginners-guide-on-logistic-regression-in-r/?utm_content=buffer47fd7&utm_medium=social&utm_source=linkedin.com&utm_campaign=buffer

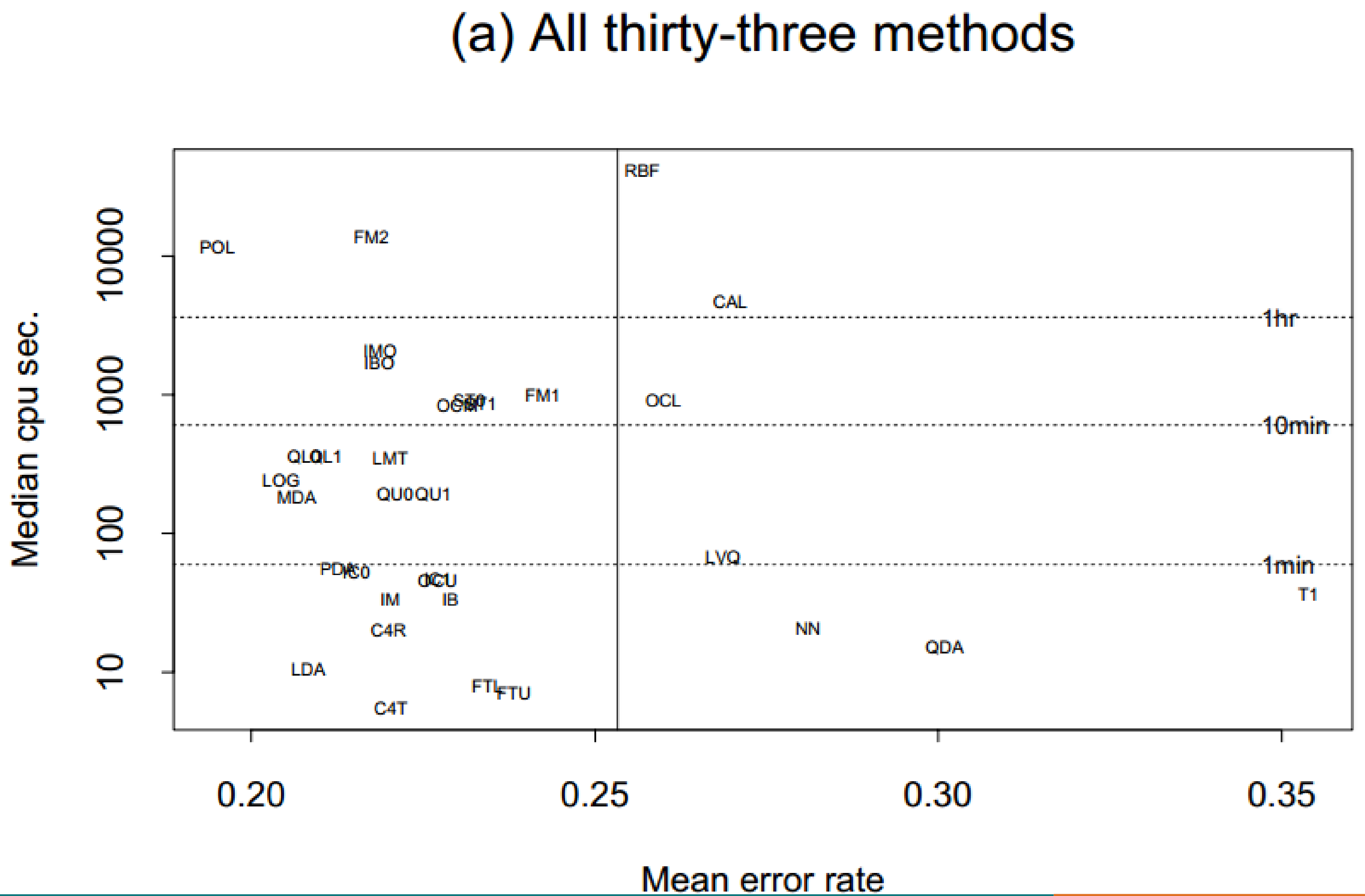
Comparison on 115 binary datasets
Source # 1

alg	mean rank	count
rf	3.04	36
svmRadial	3.36	24
gbm	3.41	25
nnet	4.88	13
rknn	5.04	10
svmPoly	5.14	11
knn	5.32	11
svmLinear	6.15	13
glmnet	6.16	15
elm	6.55	5
lvq	6.96	8
sda	7.05	5
nb	8.23	7
bst	9.08	4



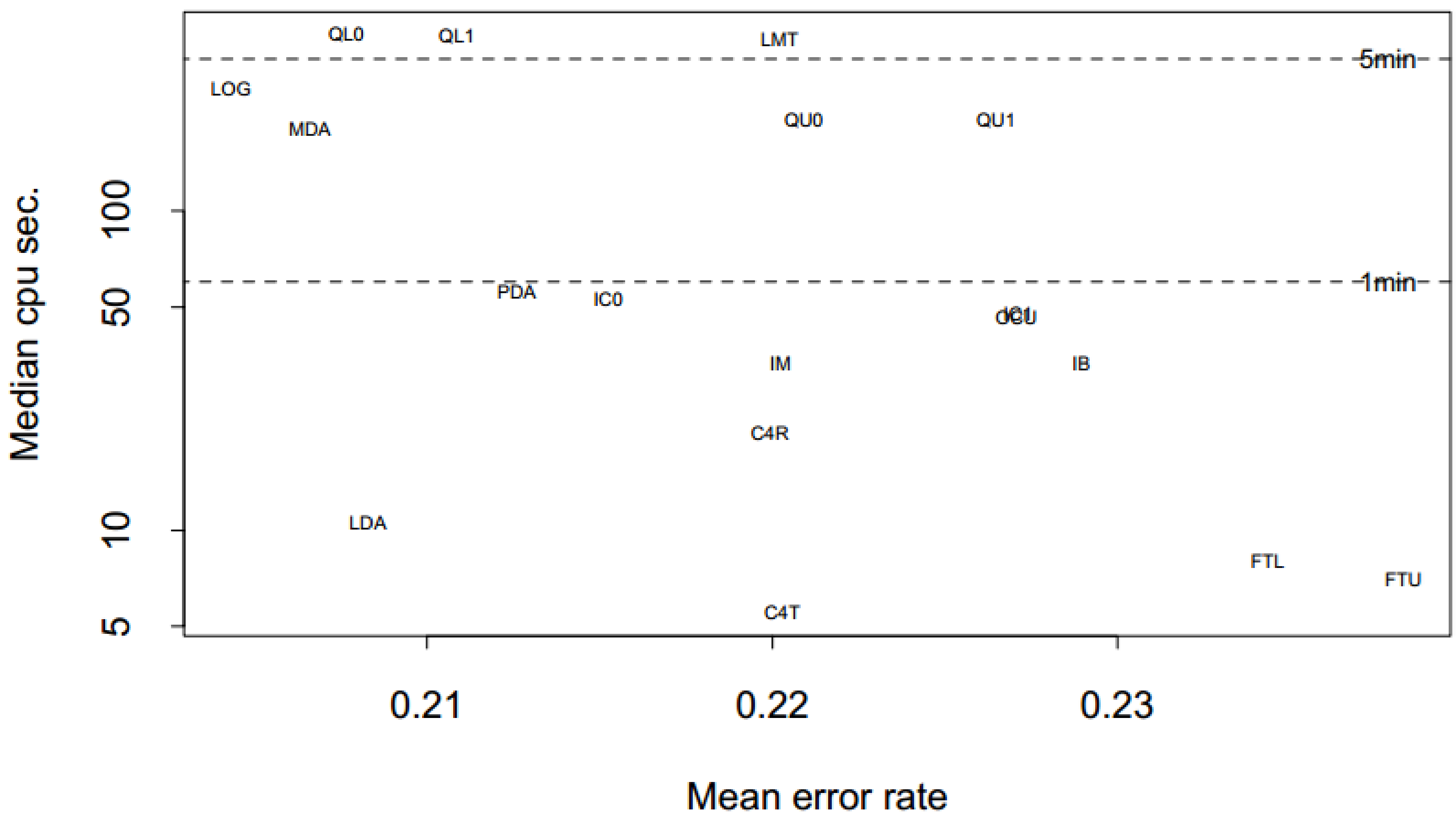
- **bst** Boosting of linear classifiers.
- **elm** Extreme learning machines
- **gbm** Gradient boosting machines.
- **glmnet** Elastic net logistic regression classifier.
- **knn** k-nearest neighbors classifier.
- **lvq** Learning vector quantization.
- **nb** Naive Bayes classifier
- **nnet** A 1-hidden layer neural network with sigmoid transfer function.
- **rf** Random forest.
- **rknn** A bagging of knn classifiers on a random subset of the original features.
- **sda** A L1 regularized linear discriminant classifier.
- **svmLinear** A SVM with linear kernel.
- **svmPoly** A SVM with polynomial kernel.
- **svmRadial** A SVM with RBF kernel.

Comparison of Classification Techniques from Source #2



Comparison of Classification Techniques from zoom on Previous Slide

(b) Less than 10min., accuracy not sig. different from POL



Principal Component Analysis

Goals

- Restructure interrelated variables
- Simplify description
- Reduce Dimensionality
- Avoid multi-collinearity problems in regression

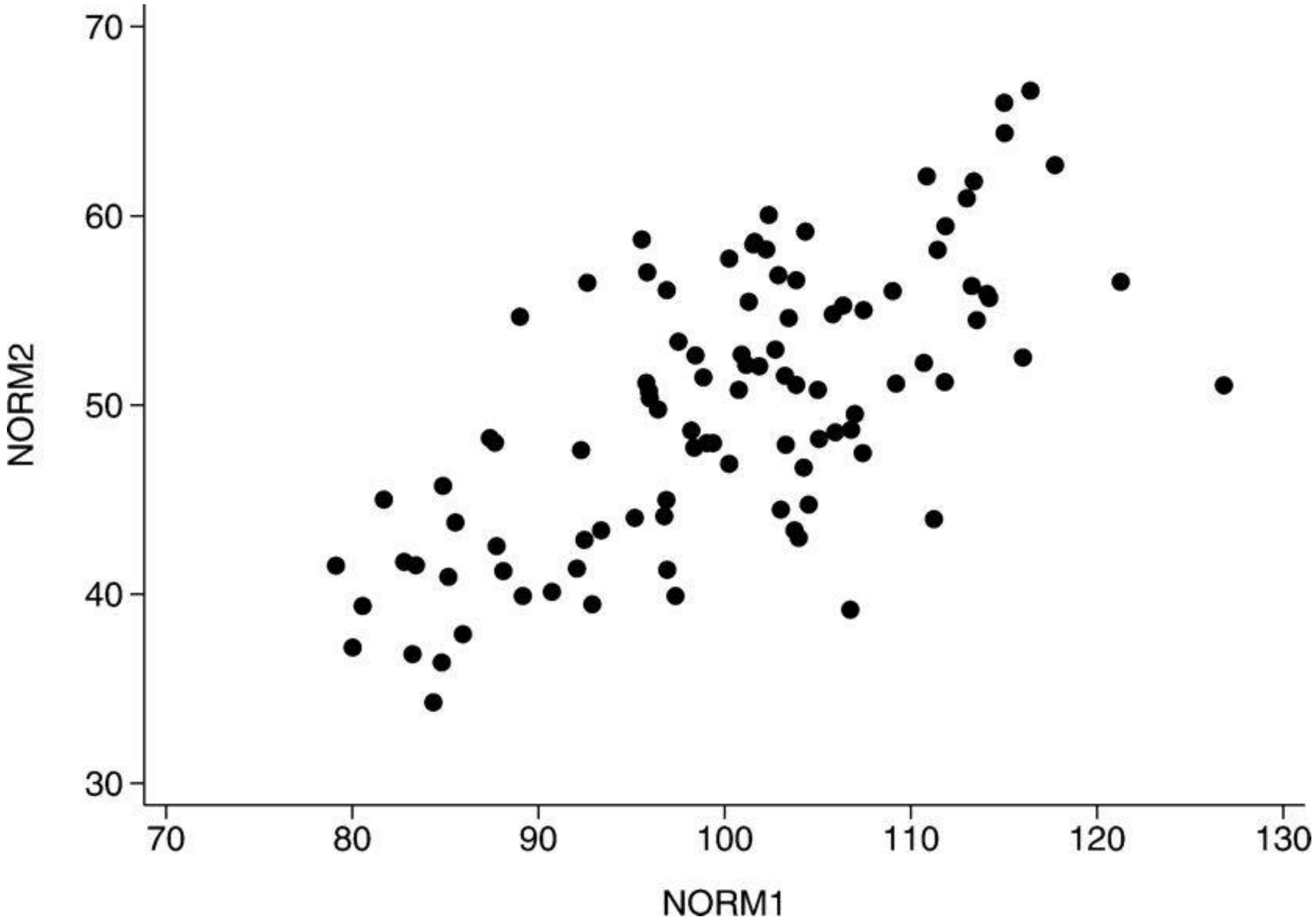
Basic Idea

- X_1 , X_2 – correlated

Transform them into:

- C_1 , C_2 – uncorrelated

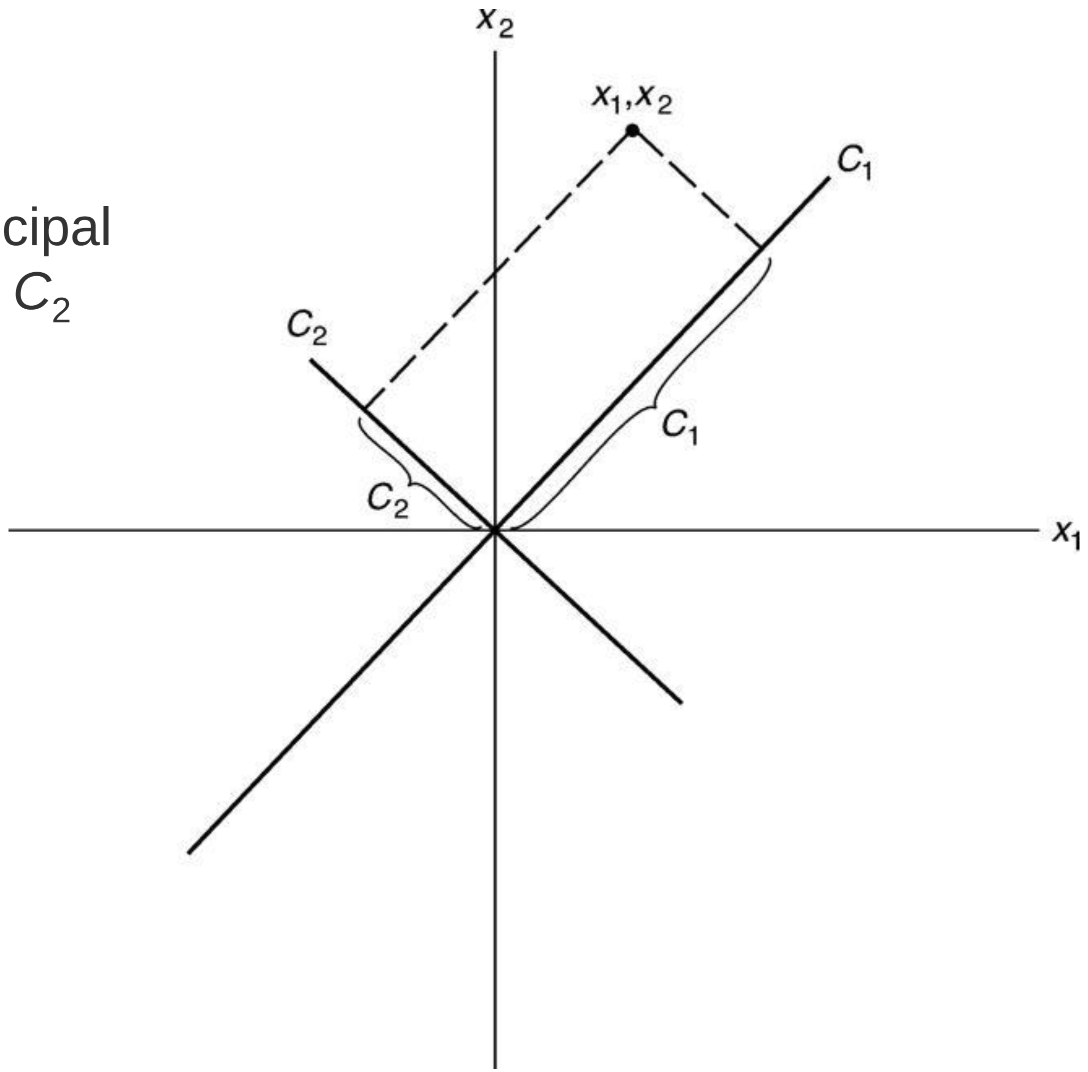
Geometric Concept (p 359)



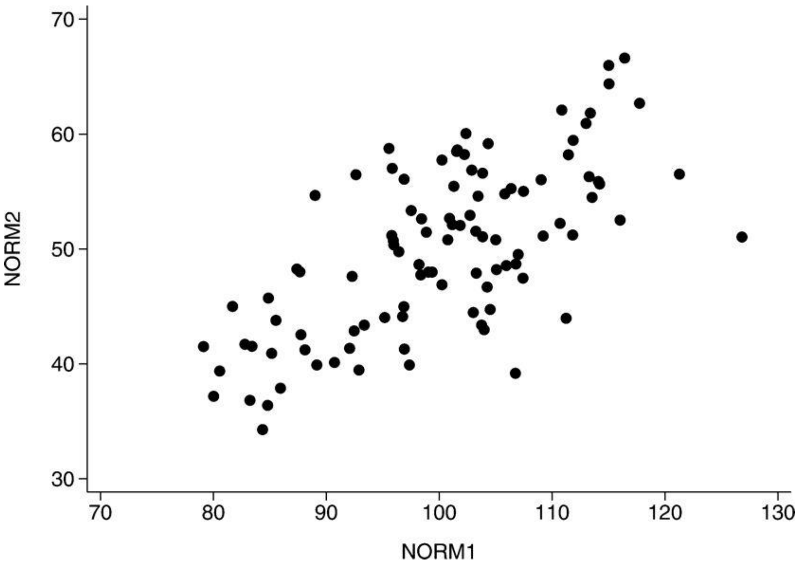
Scatter Diagram of NORM1 versus NORM2

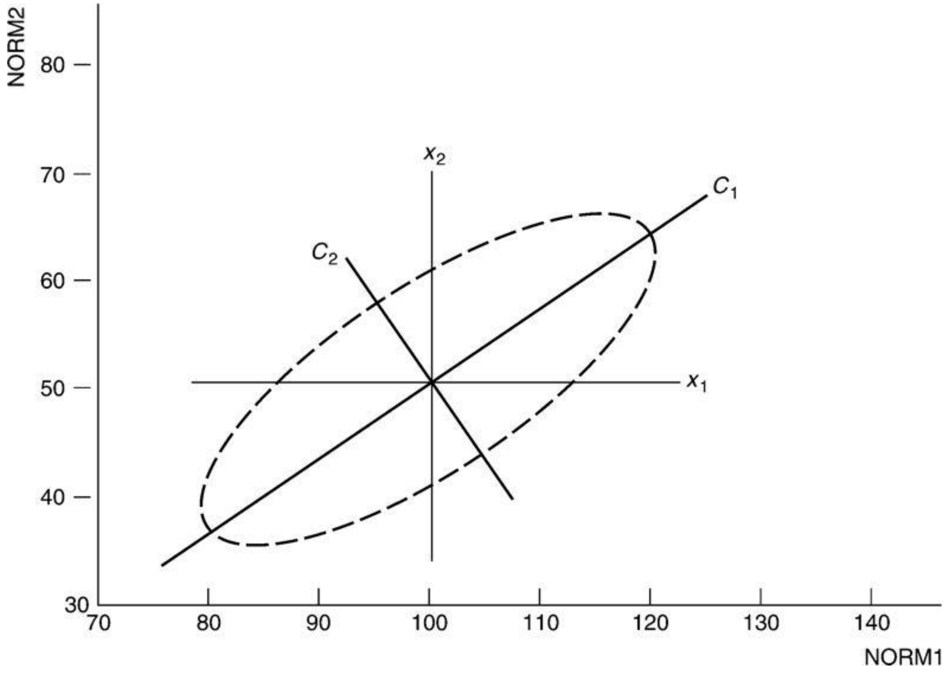
2 Principal Components: C_1 , C_2 (p 361)

Illustration of Two Principal Components C_1 and C_2



Plot of Principal Components: C_1 , C_2 (p 362)



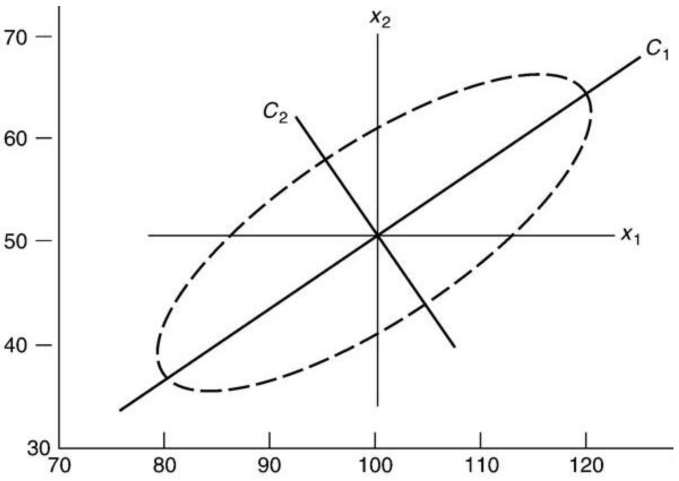


Creating new variables

- Each new variable (principal component) is a linear combination of the x variables
- For 2D case:

$$C_1 = a_{11}x_1 + a_{12}x_2$$

$$C_2 = a_{21}x_1 + a_{22}x_2$$



- New variables properties:

$$\begin{aligned}\text{mean } C_1 &= \text{mean } C_2 = 0\\ \text{Var } C_1 &= a_{11}^2 S_1^2 + a_{12}^2 S_2^2 + 2a_{11}a_{12}rS_1S_2\\ \text{Var } C_2 &= a_{21}^2 S_1^2 + a_{22}^2 S_2^2 + 2a_{21}a_{22}rS_1S_2\end{aligned}$$

Coefficients of C_1 , C_2 (p 360)

- The **coefficients** (a_{11}, a_{12}, a_{21}, a_{22}) are chosen to satisfy three requirements:

1. Var C_1 is as large as possible.
2. The N values of C_1 and C_2 are uncorrelated.
3. $a_{11}^2 + a_{12}^2 = a_{21}^2 + a_{22}^2 = 1$

Equations for Principal Components: C_1 , C_2

- $C_1 = 0.85X_1 + 0.53X_2$
- $C_2 = -0.53X_1 + 0.85X_2$
- Var C_1 = 135.0
- Var C_2 = 22.5

- Note: Dot Product of Cs: $C_1 \cdot C_2 = 0$
  - Hence they are orthogonal, therefore uncorrelated, and Var $C_1 >$ Var C_2

Principal Component Model

Principal Components: C_1 , …, C_p

- From P original variables: X_1 ,…, X_p derive P principal components C_1 ,…, C_p
- Each C_j is a linear combination of the X_i 's
- $C_j = a_{j1} X_1 + a_{j2} X_2 + \ldots + a_{jp} X_p$

Matrix Equation

$$
\begin{vmatrix} C_1 \\ C_2 \\ \vdots \\ C_p \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ & \vdots & & \\ a_{p1} & a_{p2} & \cdots & a_{pp} \end{vmatrix} \begin{vmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{vmatrix}
$$

where $\sum_{j} {a_{ij}}^2 = 1$

Properties of Principal Components:

- **Coefficients are chosen to satisfy:**

$$\text{Var } C_1 \geq \text{Var } C_2 \geq \quad . \quad . \quad . \geq \text{Var } C_p$$

- **Variance is a measure of information:**
  - **For Example, In Prostate Cancer:**
    - Gender has 0 variance, no information
    - Size of tumor has variance > 0, useful information
- **Any two principal components are orthogonal, hence uncorrelated**

Calculation of Principal Components

- **Let S be the Covariance Matrix of the X variables.**

$$S = \begin{vmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ & \vdots & & \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{vmatrix}$$

- **Then a_{ij} 's are the solution to the equation:**

$$(S - \lambda I)a = 0$$

The mathematical solution for the coefficients was originally derived by **Hotelling (1933)**

Recall Some Terminology

- Solutions to $\mathbf{(S - \lambda I)a = 0}$ are:
  - $\boldsymbol{\lambda}$ a scalar known as the eigenvalue
  - **a** a vector known as the eigenvector
- **a** is not unique. There are an infinite number of possibilities, so:
  - Choose **a** such that the sum of the squares of coefficients for each eigenvector is = 1.
  - This yields: P unique eigenvalues and P corresponding eigenvectors.

Then

- Principal Components are the **eigenvectors,**
- and their variances are the **eigenvalues**
- of the covariance matrix **S** of the **X**'s

- Variances of the $\mathbf{C}_j$ **'**s add to the sum of the variances of the original variables (total variances)

Example Revisited (p 361)

- Var X_1 = 104.0   Var X_2 = 53.5   sum = 157.5
- Var C_1 = 135.0   Var C_2 = 22.5   sum = 157.5

- Total Variance is Preserved

Transforming coefficients to correlations

- The coefficient a_{ij} can be transformed into a correlation between x_j and C_i

$$r_{ij} = \frac{a_{ij}(\text{Var } C_i)^{1/2}}{(\text{Var } x_j)^{1/2}}$$

- The first principal component: $C_1 = 0.851x_1 + 0.525x_2$
- correlation between C_1 and x_1: $r_{11} = \frac{0.85(135.04)^{1/2}}{(103.98)^{1/2}} = 0.969$
- correlation between C_1 and x_2: $r_{12} = \frac{0.525(135.04)^{1/2}}{(53.51)^{1/2}} = 0.834$

Work thru PCA by hand-1

Data Matrix	
72	176
60	115
65	180
74	160
67	150

Correlation

Corr X1:X2
0.629165

Sample Variance

Var(X1)	Var(X2)
31.3	677.2

Correlation Matrix	
1	0.629165
0.629165	1

Total variance
708.5

$$(S - \lambda \mathrm{I})a = 0$$

Eigenvalues
1.629165
0.370835

Eigenvectors	1	2
a1	0.707107	0.707107
a2	0.707107	-0.70711

Work thru PCA by hand-2

Data×Eigenvectors	
175.3625	-73.5391
123.7437	-38.8909
173.2412	-81.3173
165.463	-60.8112
153.4422	-58.6899

Sample Variance

Var(C1)	Var(C2)
445.85	262.65

Total variance
708.5

Checks:

Orthogonal?
$a_{21} \times a_{22} + a_{11} \times a_{12} = 1.11 \times 10^{-16}$

check on eigenvalue	
$s_{11} \times a_{11} + s_{12} \times a_{21} = 1.15199$	$1.15199/a_{11} = 1.629165 = 1+R$
$s_{21} \times a_{12} + s_{22} \times a_{22} = -0.26222$	$1.15199/a_{12} = -0.37083 = 1-R$

Using standardized variables

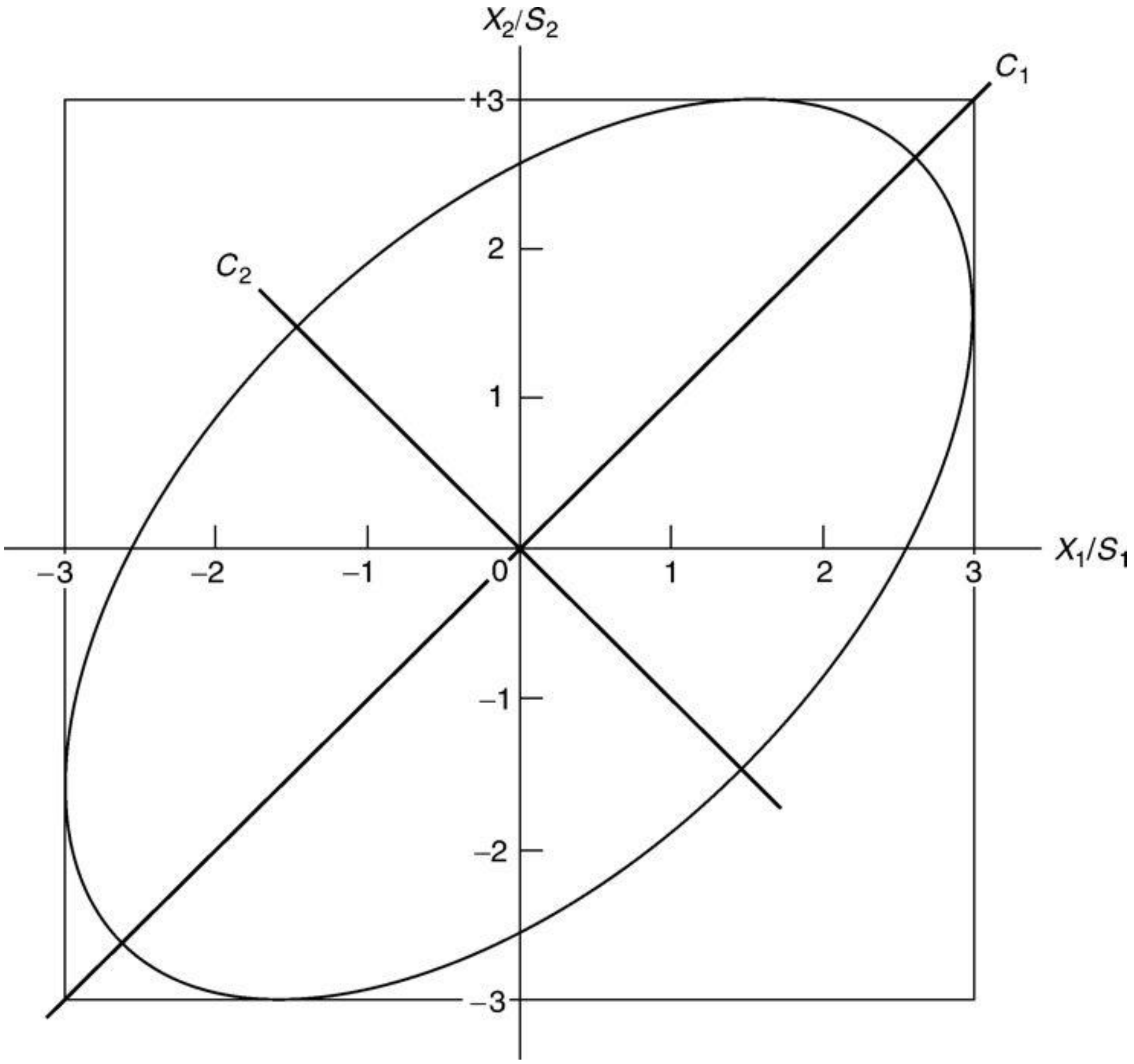
- Usually standardize:
  - Transform X $\longrightarrow$ $Z = X/SD(X)$
- Easier interpretation:
  - $\text{Var}(Z_i) = 1$
  - Total variance = P
  - $\text{Covar}(Z_i, Z_j) = r_{ij}$
  - Proportion explained by C_i = eigenvalue$_i$ / P
  - Correlation between C_i and Z_j : $r_{ij} = a_{ij}(\text{Var } C_i)^{1/2}$
  - a_{ij} : relative degree of dependence of C_i and Z_j

P 366

Principal Components for Two Standardized Variables

$$C_1 = 0.707\,\frac{\text{NORM1}}{S_1} + 0.707\,\frac{\text{NORM2}}{S_2}$$

$$C_2 = 0.707\,\frac{\text{NORM1}}{S_1} - 0.707\,\frac{\text{NORM2}}{S_2}$$



Analyzing Correlation Matrix

- Covariance matrix (S) => correlation matrix (R)
- Covariance: $cov(X,Y) = E[X-E(X)]\times[Y-E(Y)]$
- Correlation (assuming variances are positive):

  $cor(X,Y) = [cov(X,Y)]/[sd(X)\times sd(Y)]$

  $cor(X,Y) = cov\ \{[X-E(X)]/sd(X),\ [Y-EY]/sd(Y)\}$
- Correlation is a scaled version of covariance
- Correlation is dimensionless
- S and R give different Principal Components

Example Revisited

Eigenvalue	Covariance Matrix S	Correlation Matrix R
λ_1	**135.0 (85.7%)**	**1.68 (83.8%)**
λ_2	**22.5 (14.3%)**	**0.32 (16.2%)**
Total	**157.5 (100%)**	**2 (100%)**

NOTE: Different % variance explained by Principal Components

Meaning of a_{ij} (Standardized Variables)

- **Analyzing S:**
  - **Covariance (C_i , X_j) = a_{ij} * $\lambda_i^{1/2}$ / SD (X_j)**
- **Analyzing R:**
  - **Correlation (C_i , X_j) = a_{ij} * $\lambda_i^{1/2}$**

- **Correlations are a guide to interpreting Principal Components**

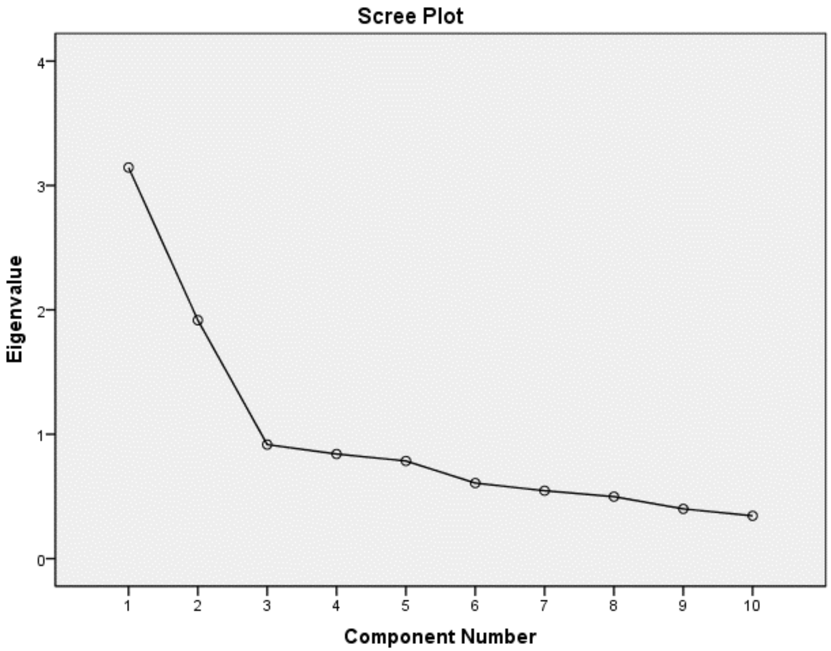
Dimension Reduction

- Retain the first m principal components as representatives of the original P variables
  - Example: keep C_1 as summary of X1, X2
  - m = 1

- Choose m large enough to explain a "large" percentage of the original total variance

Choosing m

- **Rely on existing theory and need to explain a given %**
- **Kaiser's Rule:**
  - **S: choose $\lambda_i > \text{sum of variance of X's/P}$**
  - **R: choose $\lambda_i > 1$**
- **Elbow Rule:**
  - *scree plot*



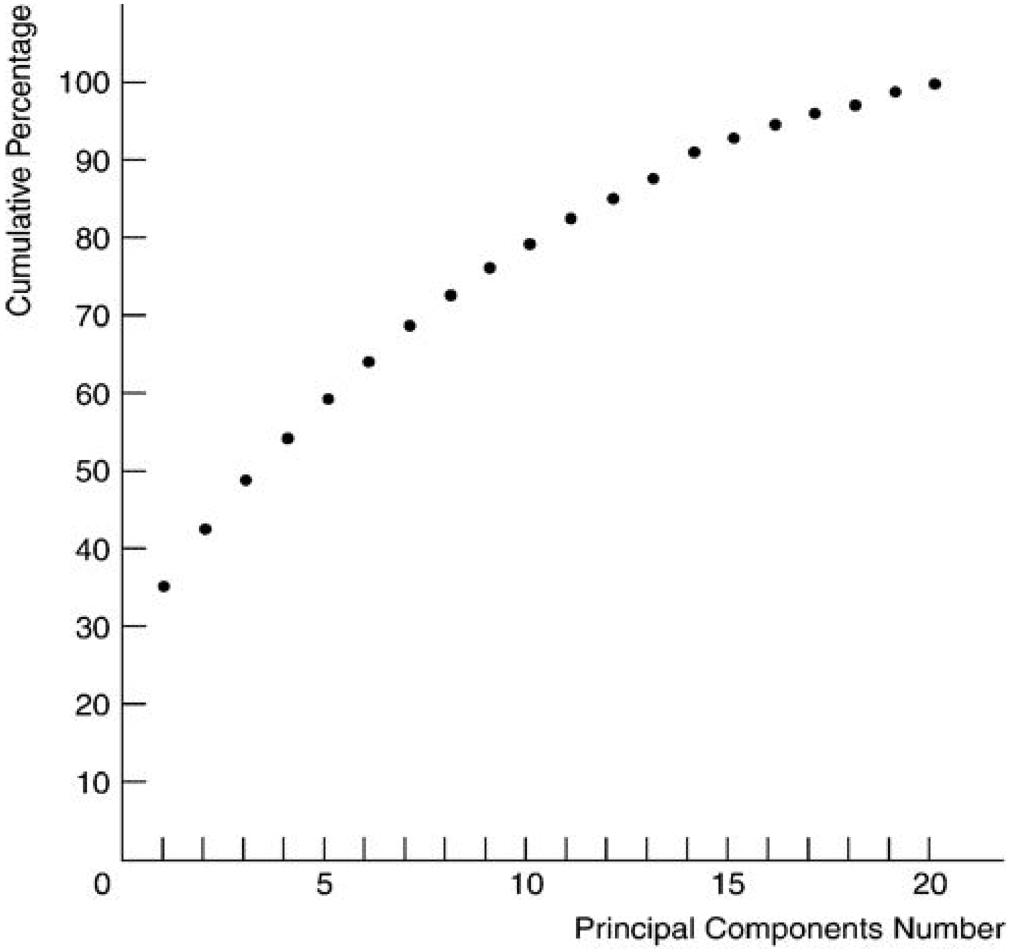
http://www.surveyanalysis.org/wiki/

Elbow Rule for Choosing m

- Start with the scree plot
- Choose a cutoff point where:
  - Lines joining consecutive points are “steep” left of the cutoff point, and
  - “flat” right of the cutoff point
- The point where the two slopes meet is the cutoff point

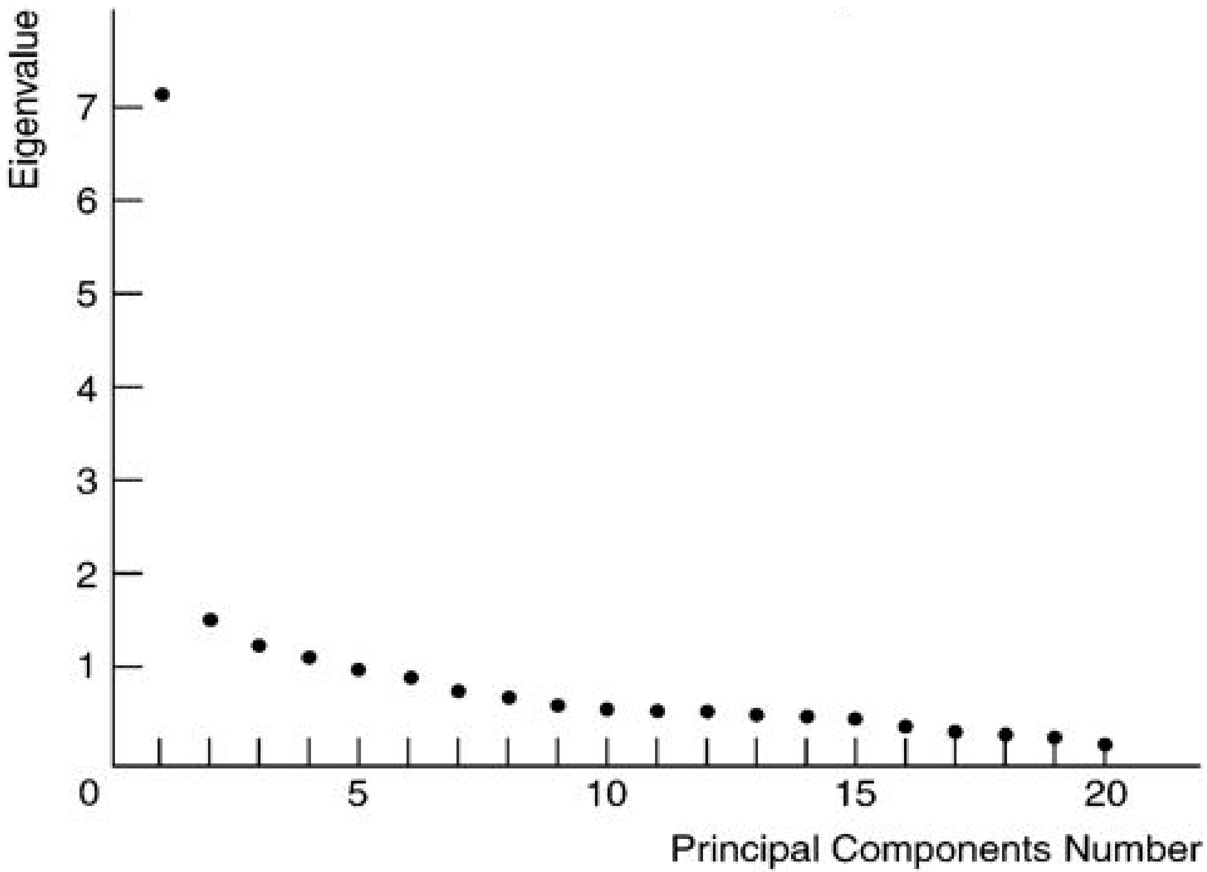
Cumulative Percentages of Total Variance for Depression Data (p 368)

Cumulative Percentages of Total Variance for Depression Data



Eigenvalues for Depression Data – Scree Plot (p 368)

Eigenvalues for Depression Data



A plot, in descending order of magnitude, of the eigenvalues of a [correlation matrix]

Depression, CESD, Example

- **Using R (Correlation, Not Language):**
  - **Choose Eigenvalues > 1**

- **Hence choose 5 Principal Components**

Principal Components for standardized CESD scale items (p 369)

Item	Principal component 1	2	3	4	5
Negative affect					
I felt that I could not shake off the blues even with the help of my family or friends.	<u>0.2774</u>	0.1450	0.0577	−0.0027	0.0883
I felt depressed.	<u>0.3132</u>	−0.0271	0.0316	0.2478	0.0244
I felt lonely.	<u>0.2678</u>	0.1547	0.0346	0.2472	−0.2183
I had crying spells.	<u>0.2436</u>	0.3194	0.1769	−0.0716	−0.1729
I felt sad.	<u>0.2868</u>	0.0497	0.1384	0.2794	−0.0411
I felt fearful.	<u>0.2206</u>	−0.0534	0.2242	0.1823	−0.3399
I thought my life had been a failure.	<u>0.2844</u>	0.1644	−0.0190	−0.0761	−0.0870
Positive affect					
I felt that I was as good as other people.	0.1081	0.3045	0.1103	<u>−0.5567</u>	−0.0976
I felt hopeful about the future.	0.1758	0.1690	−0.3962	−0.0146	<u>0.5355</u>
I was happy.	<u>0.2766</u>	0.0454	−0.0835	0.0084	0.3651
I enjoyed life.	<u>0.2433</u>	0.1048	−0.1314	0.0414	0.2419
Somatic, retarded activity					
I was bothered by things that usually don't bother me.	0.1790	−0.2300	0.1634	0.1451	0.0368
I did not feel like eating; my appetite was poor.	0.1259	−0.2126	0.2645	<u>−0.5400</u>	0.0953
I felt that everything was an effort.	0.1803	−0.4015	−0.1014	−0.2461	−0.0847
My sleep was restless.	<u>0.2004</u>	−0.2098	0.2703	0.0312	0.0834
I could not "get going."	<u>0.1924</u>	<u>−0.4174</u>	−0.1850	−0.0467	−0.0399
I had trouble keeping my mind on what I was doing.	<u>0.2097</u>	−0.3905	−0.0860	−0.0684	−0.0499
I talked less than usual.	0.1717	−0.0153	0.2019	−0.0629	0.2752
Interpersonal					
People were unfriendly.	0.1315	−0.0569	<u>−0.6326</u>	−0.0232	−0.3349
I felt that people disliked me.	<u>0.2357</u>	0.2283	−0.1932	−0.2404	−0.2909
Eigenvalues or Var C_i	7.055	1.486	1.231	1.066	1.013
Cumulative proportion	0.353	0.427	0.489	0.542	0.593
$0.5/(\mathrm{VAR}\ C_i)^{1/2}$	0.188	0.410	0.451	0.484	0.497

Reading the Output

- Here: X_1 = “I felt that I could not shake …”
  X_2 = “I felt depressed, …”

- The Principal Component are:

  $C_1 = 0.2774X_1 + 0.3132X_2 + \dots$

  $C_2 = 0.1450X_1 + 0.0271X_2 + \dots$

  etc.

Coefficients as Correlations

- **Recall: Correlation (C_i , X_j) = $a_{ij} * \lambda_i^{½}$**
- **Choose X's where Correlation > 0.5**
  - Example:
  - For C_1 , $\lambda_1 = 7.055$; $\lambda_1^{½} = 2.656$
  - Correlation > 0.5 => $a_{1j} > 0.5/2.656 = 0.188$
- **Similarly for other Principal Components**

Interpretations of Principal Components

- Loadings with r > 0.5 are underlined
- C_1 : a weighted average of most items.
  - High C_1 => respondent had many symptoms of depression. Note sign of loadings.
- C_2 : lethargy (high => energetic)
- C_3 : friendliness of others
- C_4 : self-worth / appetite
- C_5 : hopefulness

Use in Multiple Regression

- Discard the last few Principal Components, and perform regression of the remaining. This leads to more stable regression estimates.
- Alternative to variable selection
  - For example: several measures of behavior
  - Use PC_1 or PC_1 and PC_2 as summary measures of all.

Multicollinearity

- **The size of the variance of the last few Principal Components is useful as an indicator of multicollinearity among the original variables.**
- **For standardized data, the mean of C_P is zero and its variance is close to zero. So approximately:**

$$0 = a_{P1}x_1 + a_{P2}x_2 + \cdots + a_{PP}x_P$$

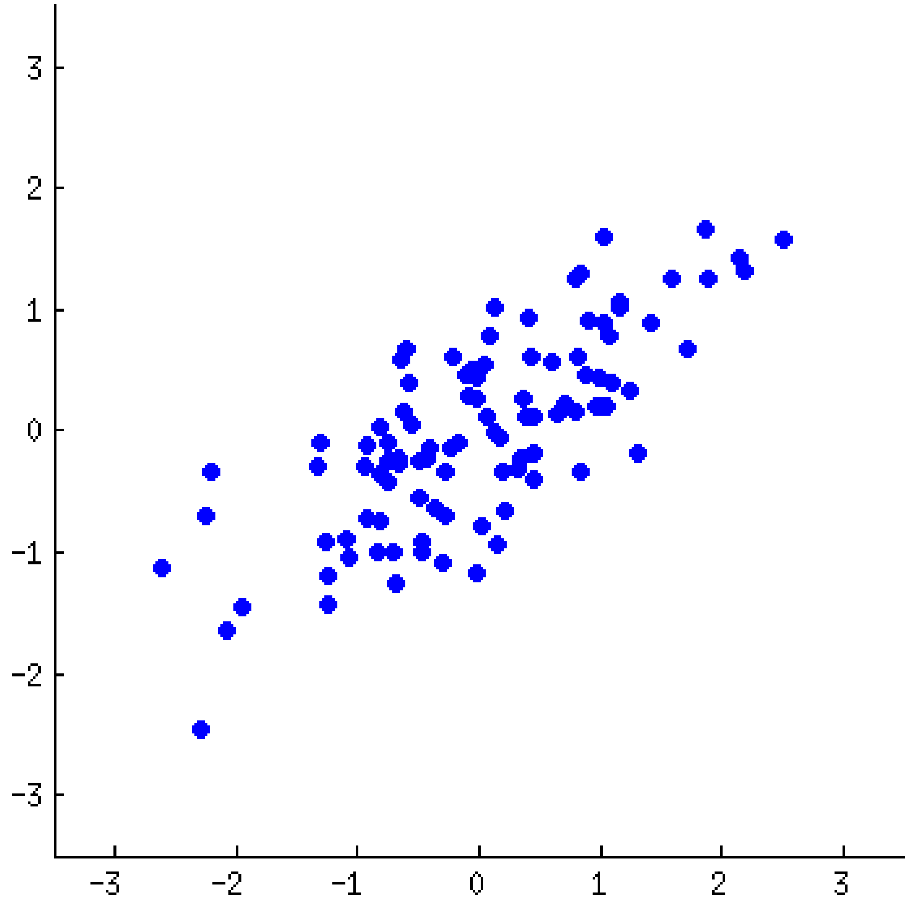
- **For example:**
  - **If $PC_{10} = .5X - .2Y - .25Z$, or $\lambda_i = 0.01 \approx 0$ then**
  - **$X \approx .4Y + .5Z$**
  - **Therefore discard X**

Summary of Computer Output

	S-PLUS/R	SAS	SPSS	Stata	STATISTICA
Covariance matrix	var	PRINCOMP	FACTOR	correlate	PCCA
Correlation matrix	cor	PRINCOMP	FACTOR	correlate	PCCA
Coefficients from raw data	princomp	PRINCOMP	FACTOR	pca	PCCA
Coefficients from standardized data	princomp	PRINCOMP	FACTOR	pca	PCCA
Correlation between xs and Cs	factanal	FACTOR	FACTOR	predict;correlate	PCCA
Eigenvalues	princomp	PRINCOMP	FACTOR	pca	PCCA
Cum. proportion of variance explained	princomp	PRINCOMP	FACTOR	pca	PCCA
Principal component scores saved	princomp	PRINCOMP	FACTOR	predict	PCCA
Plot of principal components	biplot	GPLOT	FACTOR	loadingplot	PCCA
Scree plots	screeplot	FACTOR	FACTOR	screeplot	PCCA

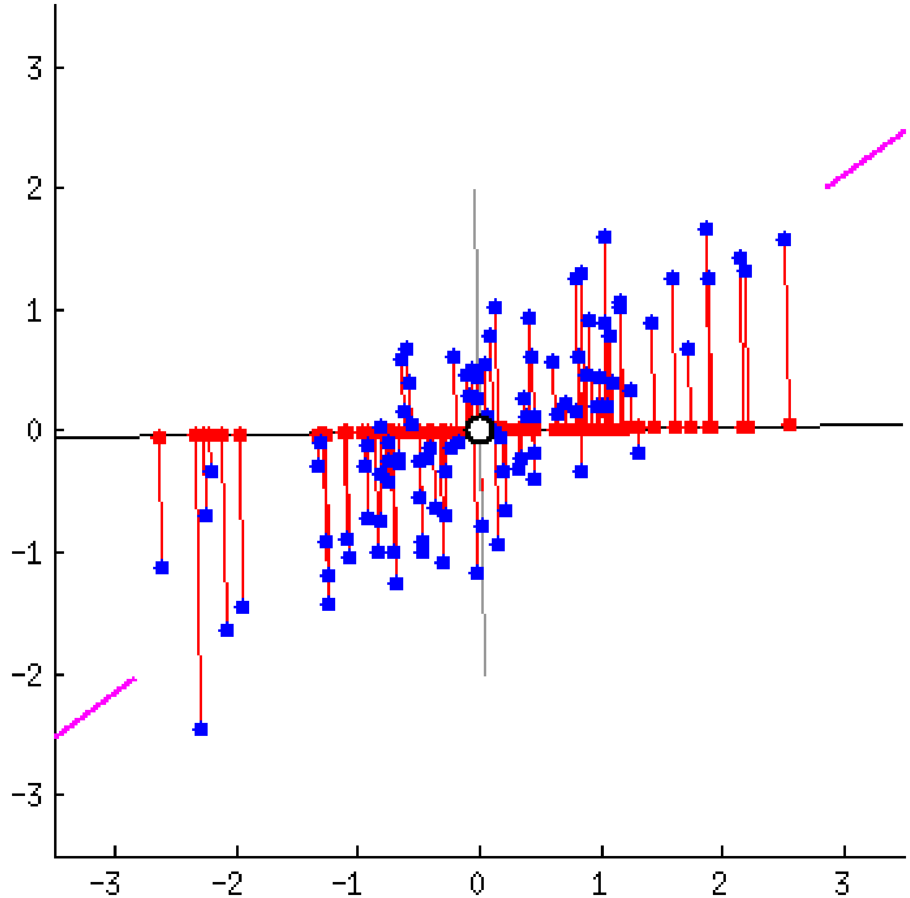
Caveats

- Principal Components derived from standardized variables differ from those derived from original variables.
- Interpretation is easier if data arise from, or are transformed to have, a symmetric distribution
- It is important that measurements are accurate, especially for detection for collinearity

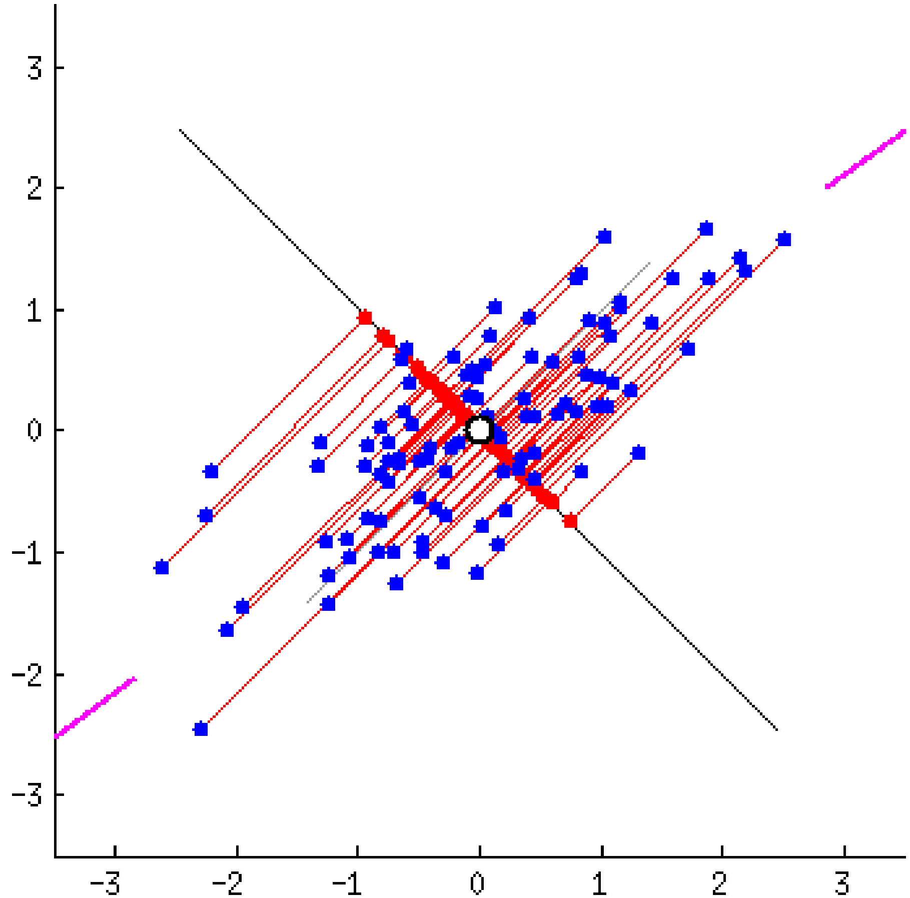


https://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues

Projections for Different Lines



https://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues



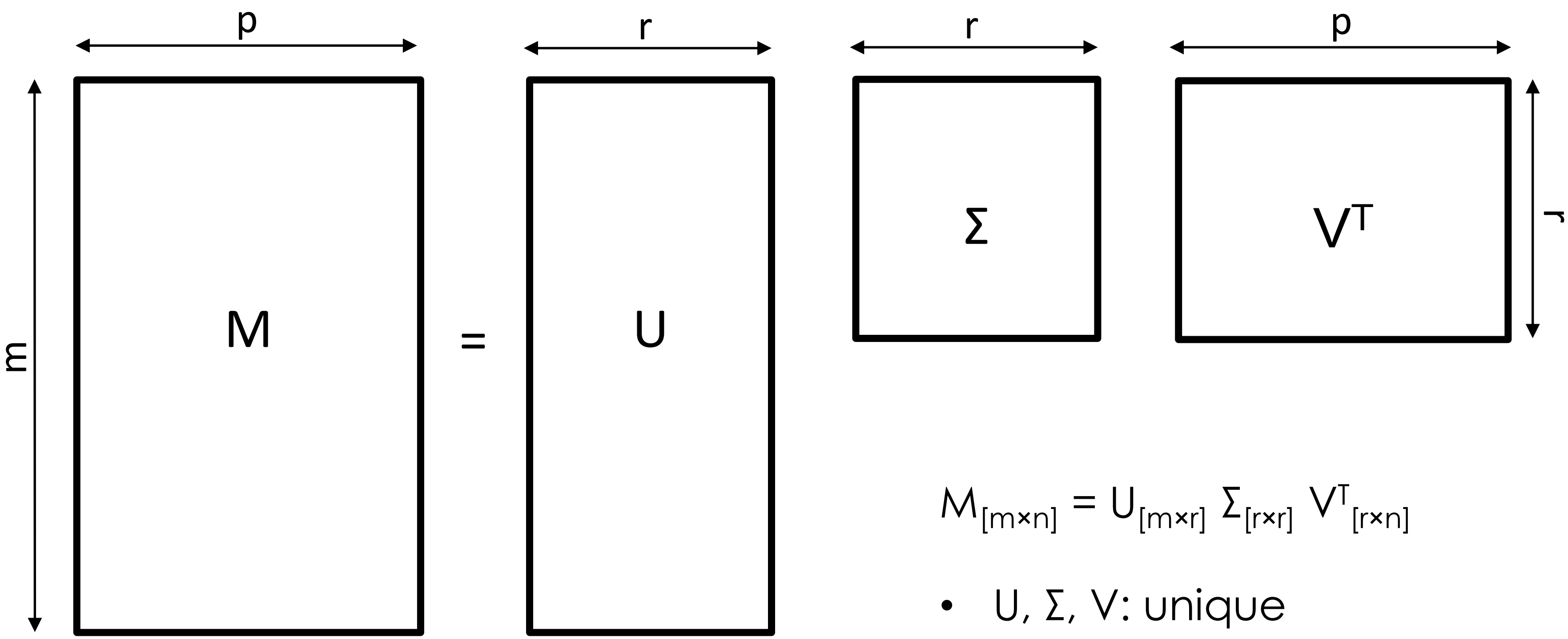
https://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues

Singular Value Decomposition

Singular Value Decomposition (SVD)

- SVD is a factorization of a matrix: $M = U \Sigma V^T$
- M is the “ratings” matrix
- U and V are the eigenvectors of $M M^T$ and $M^T M$ respectively.
- Σ is a diagonal matrix of singular values (think of them as weighting factors for U and V).
- The singular values are the square roots of the eigenvalues of $M M^T$ corresponding to the eigenvectors in U and V.
- $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p$

Singular Value Decomposition (SVD)



$$M_{[m\times n]} = U_{[m\times r]}\, \Sigma_{[r\times r]}\, V^T_{[r\times n]}$$

- U, Σ, V: unique
- U, V: column orthogonal
- Σ: diagonal

Singular Value Decomposition (SVD)

Min sum squares of projection errors

$(U\Sigma)^T$: Projection Axis

Singular Value Decomposition (SVD) and Principal Component Analysis (PCA)

- Recall PCA creates coordinates with greatest influence on the analysis.
- It does this by calculating the eigenvalues and eigenvectors of the covariance (or correlation) matrix.
- One can calculate SVD using PCA in a variety of ways, for example:
  - Translate each data column to have mean=0
  - Then diagonalizing $X^T X$ provides the principal components.
- At least as likely to calculate PCA using SVD at scale. For example:
  - Subtract the column means from the data matrix.
  - Take SVD
  - The Columns of V_K are principal components.
  - Values of the w_l, i.e. diagonal elements of Σ give “importance” of each element.

SVD Recommender Example - 1
X = User/Item

Input matrix:

5.000	5.000	1.400	2.400
2.400	2.000	1.400	2.400
4.000	1.000	3.000	2.000
3.000	3.000	3.000	3.000
2.400	3.200	1.000	4.000

Matrix Algebra using: http://www.bluebit.gr/matrix-calculator/multiply.aspx

SVD Recommender Example - 2
SVD of X

Singular Value Decomposition:

U:

-0.581	-0.395	-0.710	0.014
-0.329	0.014	0.189	-0.189
-0.391	0.791	-0.111	-0.404
-0.466	0.206	0.287	0.809
-0.430	-0.420	0.604	-0.383

S:

12.631	0.000	0.000	0.000
0.000	2.922	0.000	0.000
0.000	0.000	2.275	0.000
0.000	0.000	0.000	0.968

V^T

-0.609	-0.532	-0.338	-0.481
0.284	-0.645	0.697	-0.136
-0.540	-0.214	0.177	0.794
-0.507	0.505	0.607	-0.344

SVD Recommender Example - 3

U

Input matrix A:

-0.581	-0.395	-0.710
-0.329	0.014	0.189
-0.391	0.791	-0.111
-0.466	0.206	0.287
-0.430	-0.420	0.604

S

Input matrix B:

12.631	0.000	0.000
0.000	2.922	0.000
0.000	0.000	2.275

UxS

Matrix product A*B

-7.339	-1.154	-1.615
-4.156	0.041	0.430
-4.939	2.311	-0.253
-5.886	0.602	0.653
-5.431	-1.227	1.374

UxS

Input matrix A:

-7.339	-1.154	-1.615
-4.156	0.041	0.430
-4.939	2.311	-0.253
-5.886	0.602	0.653
-5.431	-1.227	1.374

V^T

Input matrix B:

-0.609	-0.532	-0.338	-0.481
0.284	-0.645	0.697	-0.136
-0.540	-0.214	0.177	0.794

UxS xV^T

Matrix product A*B

5.014	4.994	1.390	2.405
2.310	2.093	1.509	2.335
3.801	1.191	3.235	1.860
3.403	2.603	2.525	3.268
2.217	3.387	1.224	3.870

Dimension Reduction
(aka Matrix Factorization Techniques)
PCA and SVD

PCA: Principle Components Analysis.

SVD: Singular Value Decomposition.

SVD and PCA are closely related.

Why we use SVD and PCA?

- A powerful tool for analyzing data and finding patterns.
- Used for compression. So you can reduce the number of dimensions without much loss of information.

PCA and SVD

Summary for PCA and SVD

Objective: find the principal components P of a data matrix A(n,m).

1. First zero mean the columns of A (translate the origin to the center of gravity).
2. Apply PCA or SVD to find the principle components (P) of A.

PCA:

I. Calculate the covariance matrix C= $\frac{A\,A^T}{n}$

II. p = the eigenvectors of C.

III. The variances in each new dimension is given by the eigenvalues.

SVD:

I. Calculate the SVD of A.

II. P = V: the right singular vectors.

III. The variances are given by the squaring the singular values.

3. Project the data onto the feature space. F = P $\times$ A
4. Optional: Reconstruct A from Y where A'is the compressed version of A.

Comparison between PCA and SVD

- They both are eigenvalue/eigenvector methods

- PCA is faster than SVD when:

the number of observations > the number of variables

- PCA uses SVD in its calculation

- Geometrically PCA corresponds to "centering the dataset", and then rotating it to align the axis of highest variance with the principle axis.

Ref.: https://www.quora.com/What-is-the-difference-between-PCA-and-SVD

A Bit of Review

- Preparation of Data: Outliers, Null Values, etc.
- Regression – Simple and Multiple
- Discriminant Analysis
- Logistic Regression
- Cluster Analysis – Hierarchical, K-Means, Density
- Dimension Reduction
  - Principal Component Analysis
  - Singular Value Decomposition
  - Factor Analysis

Project Presentation

- The project will be presented in the order scheduled.
- Please find the template for your presentation here
- You may use your laptop to make a presentation.
- The paper presentations will last 15 minutes and then there will be 5 minutes for questions.
- Presentations will be strictly timed, just as if being done at a professional conference.
- You are advised to practice your presentation before giving it, with the overhead materials you will actually use, to make sure it does not take too long. Don't try to memorize the presentation -- use the overheads as clues about what you should be talking about, but also (very important) don't just READ the overheads.
- All group members should participate in the presentation. For example, if you are in a group of three, each of you should present about five minutes.

Term Project

- Report (both PDF and Word files)
- Presentation (PPT file)
- Code

You need to submit them up to Dec 4, 2018

The presentation and code only need to be submitted! and your project credit comes from how you present the project and how is the final written report. Report needs to be complete and you should not mention “it is mentioned/presented in the code or presentation”. All points should be explained and discussed in the report and do not leave any figure or table without explanations or discussion. Also, you can expand/modify the report and address the points that you missed in your presentation.

Term Paper

The report needs detail the following:

- Title
- Names
- Abstract
- Introduction
- Problem description
- Evaluation of database (no. variables, instances, etc.)
- Data processing and preparation
- Methods used (regression, classification, and clustering) and why, perhaps you tried many which should also be reported
- Dimension reduction method(s)
- Results (plots and tables), Comparisons, and Discussions
- Conclusion
- Future research
- References

Write this as if you were trying to publish results in a professional conference/journal. Your evaluation will be based on this viewpoint!

stevens.edu

Amir H Gandomi; PhD
Assistant Professor of Information Systems
a.h.gandomi@stevens.edu