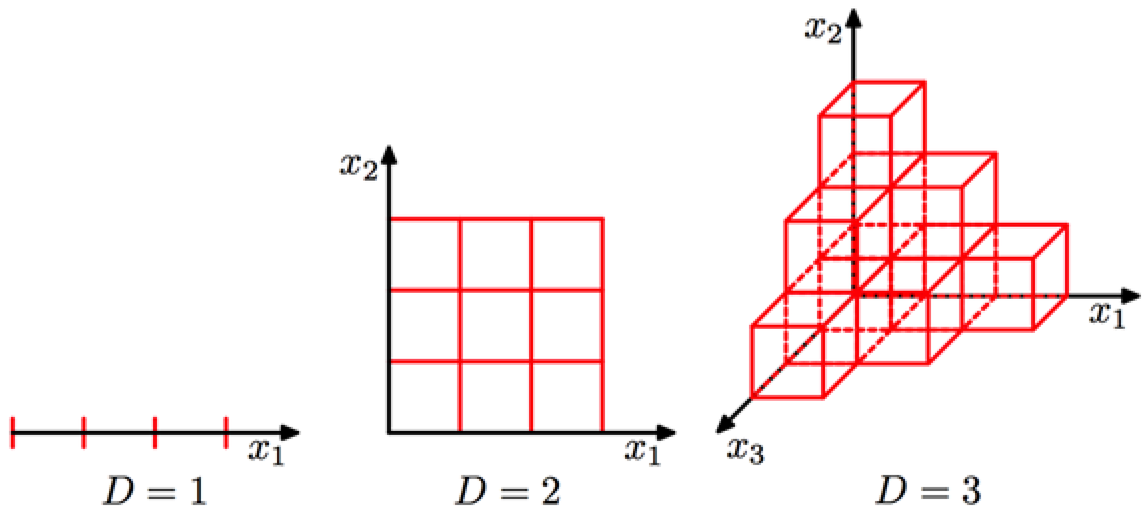
Do It Skills
do it. enjoy it.

PCA

PRINCIPAL COMPONENT ANALYSIS
DIMENSIONALITY REDUCTION TECHNIQUE

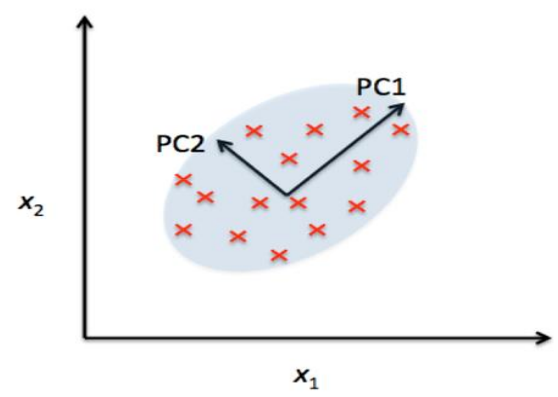
CURSE OF DIMENSIONS



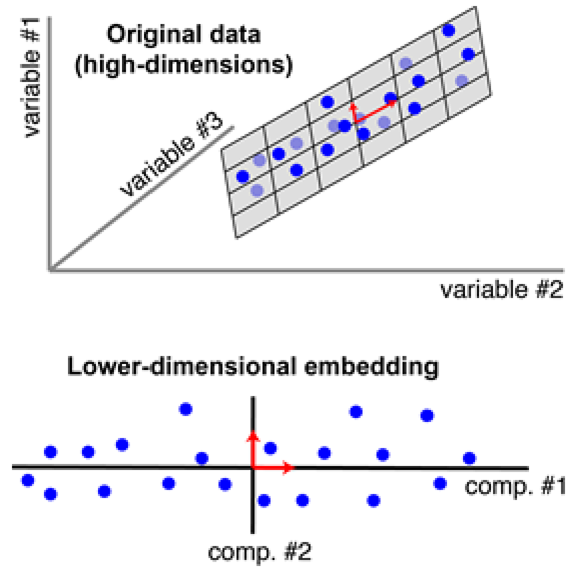
Many algorithms that work fine in low dimensions become intractable when the input is high-dimensional.

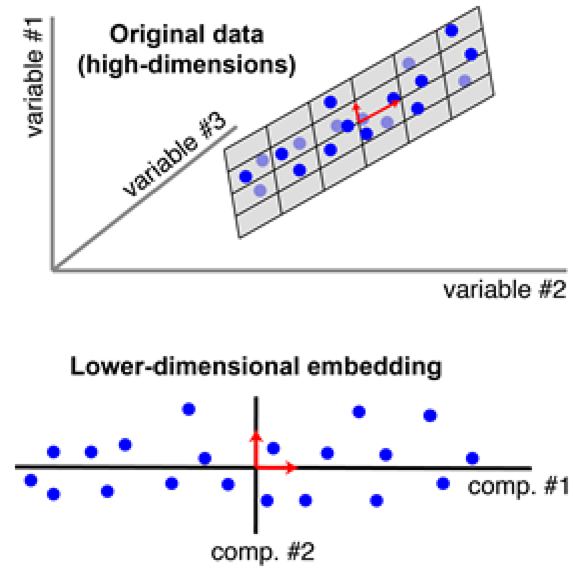
PCA

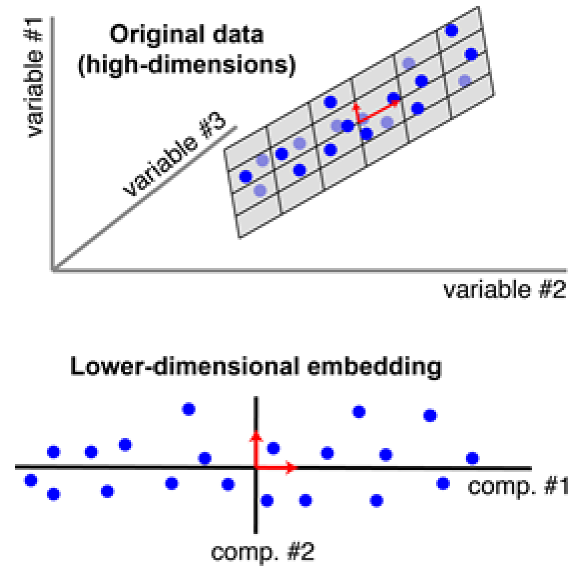
PCA is described as unsupervised



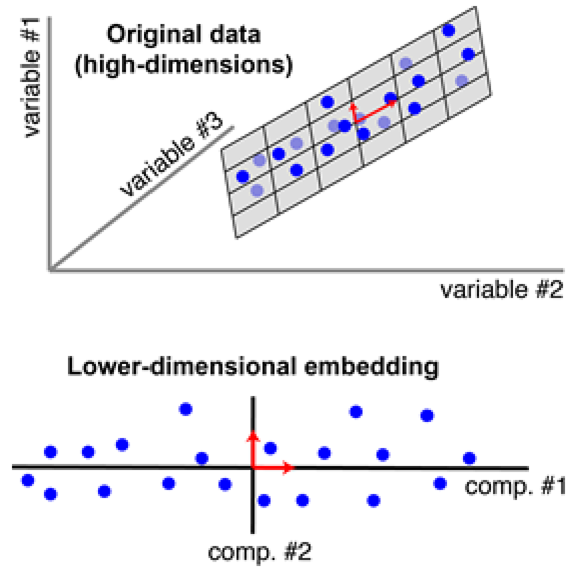
- In PCA component axes that maximize the variance
- PCA searches for the direction that have largest variations
- PCA require fewer computations
- Application of PCA in the prominent field of criminal investigation is beneficial


Original data
(high-dimensions)
variable #1
variable #3
variable #2
Lower-dimensional embedding
comp. #1
comp. #2


Original data
(high-dimensions)
variable #1
variable #2
variable #3
Lower-dimensional embedding
comp. #1
comp. #2

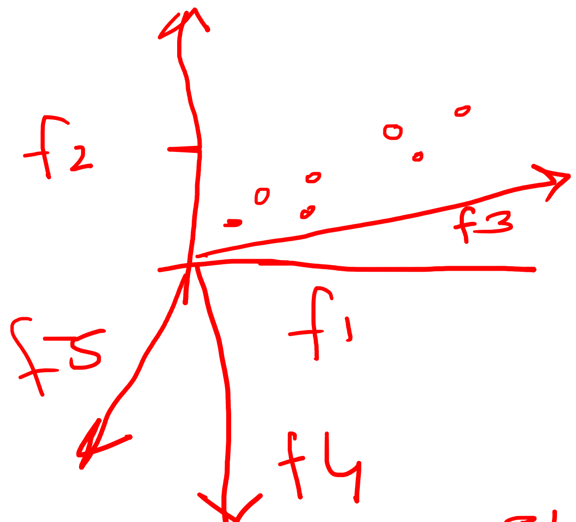


Do It Skills
do it. enjoy it.

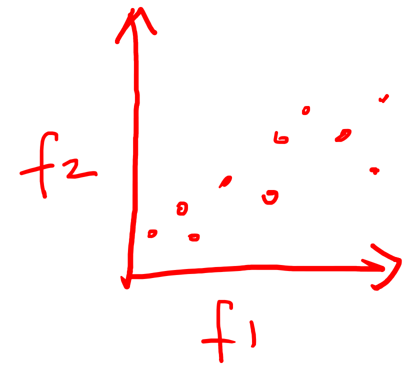

Original data
(high-dimensions)
variable #1
variable #3
variable #2
Lower-dimensional embedding
comp. #1
comp. #2

PC1 PC2 PC3 PC4

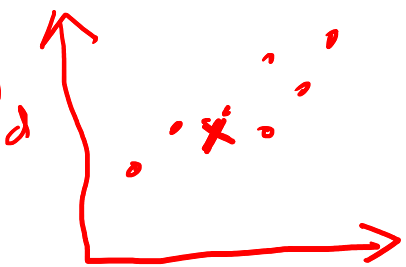
	f1	f2	f3	f4
c1	1	2	3	4
c2	5	5	6	7
c3	1	4	2	3
c4	5	3	2	1
c5	8	1	2	2



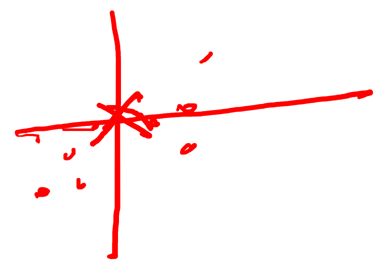
Step 1



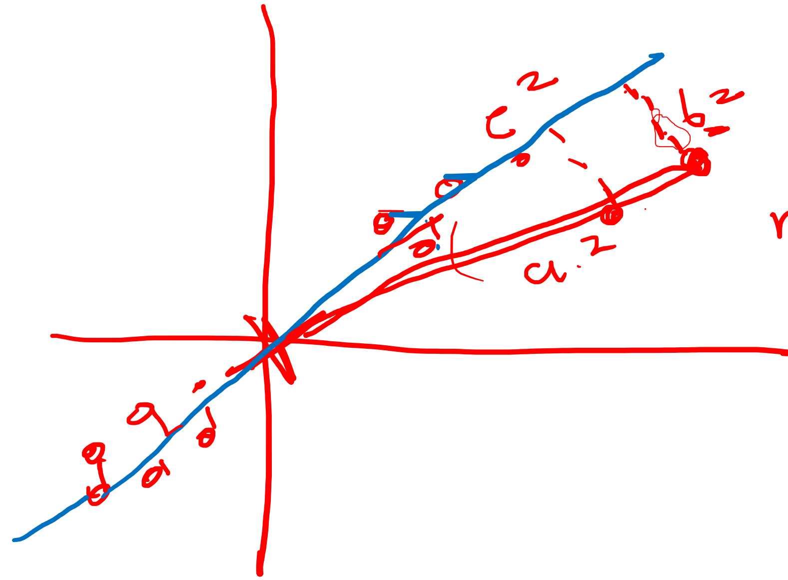
step2
mean
centroid



step 3



	f1	f2	f3	f4
c1	1	2	3	4
c2	5	5	6	7
c3	1	4	2	3
c4	5	3	2	1
c5	8	1	2	2



$a^2 = b^2 + c^2$

maximum distance

WHAT IS PCA (NEED)

- PCA is mostly used as a tool in exploratory data analysis and for making predictive models.
- Its operation can be thought of as revealing the internal structure of the data in a way that best explains the variance in the data.
- Principal component analysis (**PCA**) is a technique **used** to emphasize variation and bring out strong patterns in a dataset. It's often **used** to make data easy to explore and visualize.
- The main idea of **principal component analysis** (**PCA**) is to reduce the dimensionality of a data set consisting of many variables correlated with each other, either heavily or lightly, while retaining the variation present in the dataset, up to the maximum extent. ... As a layman, it is a method of summarizing data.

WHAT IS DATA REDUCTION

- Data reduction is summarization of data with p variables by a smaller set of (k) derived variables.
- These k derived variables are linear combinations of original variables.
- In short, n * p matrix is reduced to n * K matrix.

WHERE IS PCA USED?(APPLICATION)

- **PCA** is predominantly **used** as a dimensionality reduction technique in domains like facial recognition, computer vision and image compression. It is also **used** for finding patterns in data of high dimension in the field of finance, data mining, bioinformatics, psychology, etc.
- **PCA** is a widely used statistical tool for dimension reduction. The **objective of PCA** is to find common factors, the so–called **principal components**, in form of linear combinations of the variables
- **PCA** is a most widely used tool in exploratory data analysis and in machine learning for predictive models.
- Is PCA supervised or unsupervised?
- **Principal component analysis** (**PCA**) is an **unsupervised** technique used to preprocess and reduce the dimensionality of high-dimensional datasets while preserving the original structure and relationships inherent to the original dataset so that machine learning models can still learn from them and be used to make accurate

WHAT TYPE OF DATA SHOULD BE USED FOR PCA?

- PCA works best on data set having 3 or higher dimensions. Because, with higher dimensions, it becomes increasingly difficult to make interpretations from the resultant cloud of data. PCA is applied on a data set with numeric **variables**.

WHAT IS EIGENVALUE AND EIGENVECTOR ?

Definition : A nonzero vector x is an eigenvector(or characteristic vector) of a square matrix A if there exists a scalar λ such that Ax= λ x. Then λ is an eigenvalue (or characteristic value) of A.

The zero vector can not be an eigenvector even though A0= λ 0 but λ =0 can be eigenvalue

Example:

$$Show\ x = \begin{bmatrix} 2 \\ 1 \end{bmatrix} is\,an\,eigenvector\ for\ A = \begin{bmatrix} 2 & -4 \\ 3 & -6 \end{bmatrix}$$

$$Solution : Ax = \begin{bmatrix} 2 & -4 \\ 3 & -6 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$But\ \ for\ \ \lambda = 0,\ \ \lambda x = 0\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$Thus, x\,is\,an\,eigenvector\,of\ A, and\,\lambda = 0\ is\,an\,eigenvalue.$$

Do It Skills
do it. enjoy it.

DEFINITION

An eigenvector of an nXn matrix A is a nonzero vector x such that Ax = λx for some scalar λ. A scalar λ is called an eigenvalue of A if there is a nontrivial solution x of Ax = λx ; such an x is called an eigenvector corresponding to λ.

Eigenvector of Matrix **A**

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

Eigenvalue of Matrix **A**

USE OF EIGENVALUES AND EIGENVECTORS

- Eigenvalues and eigenvectors feature prominently in the analysis of linear transformations.
- Eigenvalues and eigenvectors have a wide range of applications, for example in stability analysis, vibration analysis, atomic orbitals, facial recognition and matrix diagonalization.
- In essence, an eigenvector v of a linear transformation *T* is a nonzero vector that, when *T* is applied to it, does not change direction. Applying *T* to the eigenvector only scales the eigenvector by the scalar value λ, called an eigenvalue. This condition can be written as the equation

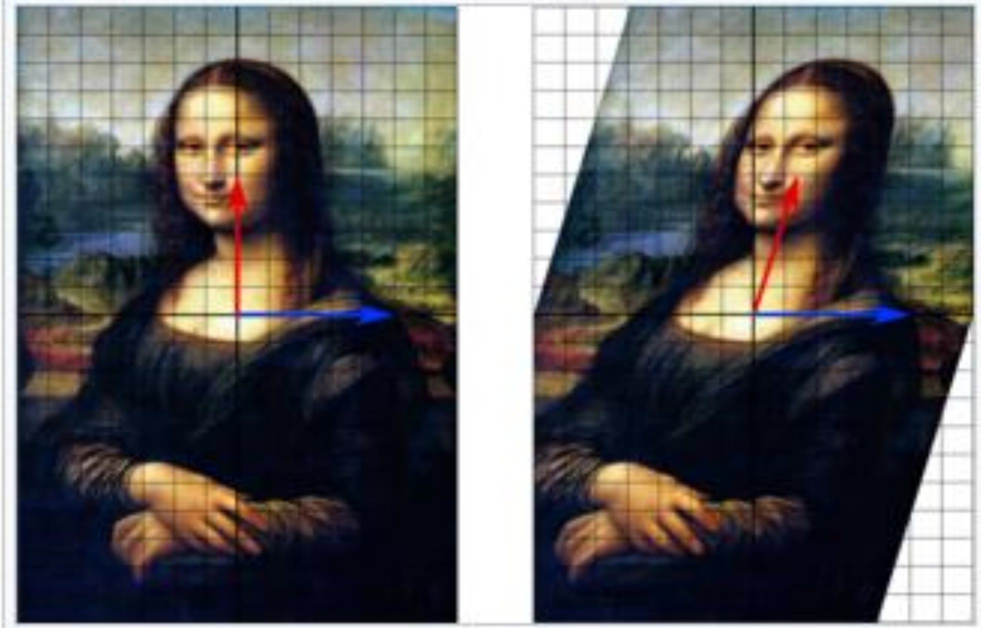
$$T(\mathbf{v}) = \lambda \mathbf{v},$$

referred to as the eigenvalue equation or eigenequation. In general, λ may be any scalar. For example, λ may be negative, in which case the eigenvector reverses direction as part of the scaling, or it may be zero or complex.

Do It Skills
do it. enjoy it.

EXAMPLE OF EIGENVALUE AND EIGENVECTOR

The Mona Lisa example pictured here provides a simple illustration. Each point on the painting can be represented as a vector pointing from the center of the painting to that point. The linear transformation in this example is called a shear mapping. Points in the top half are moved to the right and points in the bottom half are moved to the left proportional to how far they are from the horizontal axis that goes through the middle of the painting. The vectors pointing to each point in the original image are therefore tilted right or left and made longer or shorter by the transformation. Points *along* the horizontal axis do not move at all when this transformation is applied. Therefore, any vector that points directly to the right or left with no vertical component is an eigenvector of this transformation because the mapping does not change its direction. Moreover, these eigenvectors all have an eigenvalue equal to one because the mapping does not change their length, either.

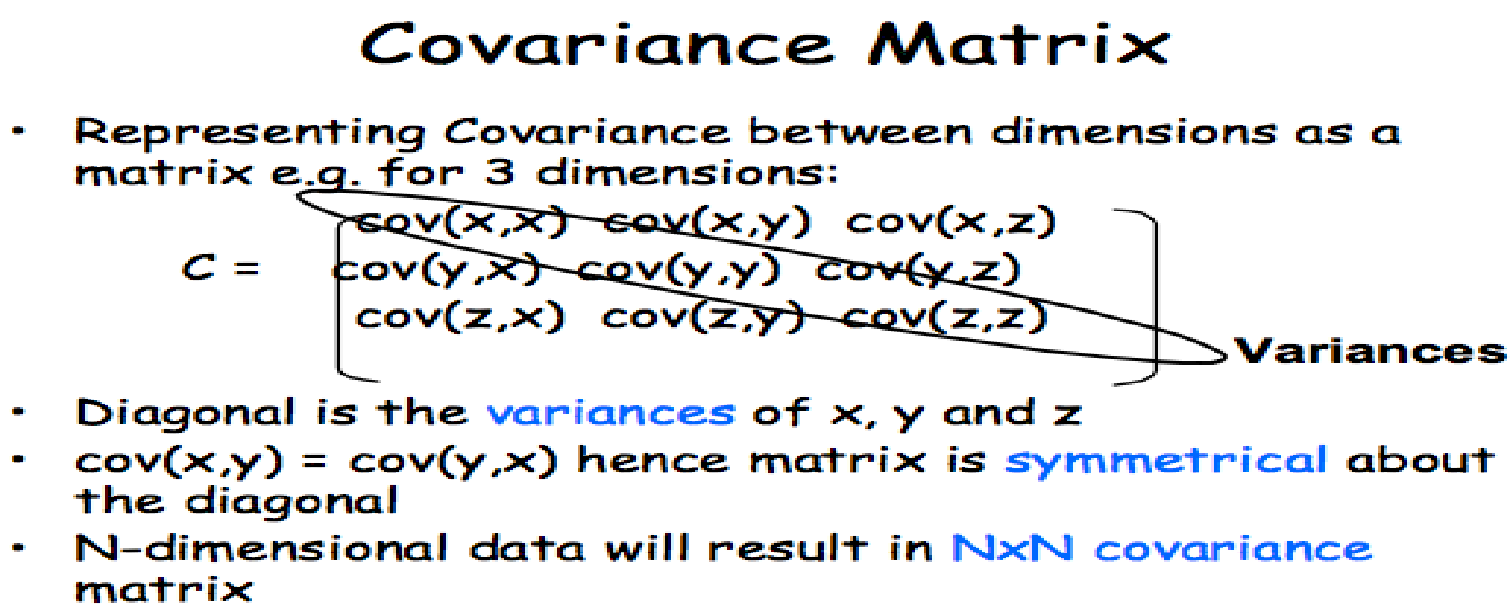


In this shear mapping the red arrow changes direction, but the blue arrow does not. The blue arrow is an eigenvector of this shear mapping because it does not change direction, and since its length is unchanged, its eigenvalue is 1.

Do It Skills
do it. enjoy it.

COVARIANCE MATRIX

- In probability theory and statistics, a covariance matrix is a square matrix giving the covariance between each pair of elements of a given random vector. In the matrix diagonal there are variances, i.e., the covariance of each element with itself.



PRINCIPAL COMPONENT ANALYSIS

- PCA uses correlation structure of original variables and derives p linear combinations which are uncorrelated.
- Each PC provides unique information about the data.
- Although 'p' principal components are derived, first 'k' principal components are expected to explain most of the variability in the data.

Do It Skills
do it. enjoy it.

PCA: GENERAL CONCEPT

- From p original variables: $x_1, x_2, \ldots, x_p$, derive p new variables $y_1, y_2, \ldots y_p$, :

$y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1p}x_p$

$y_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2p}x_p$

...

$y_p = a_{p1}x_1 + a_{p2}x_2 + \ldots + a_{pp}x_p$

Note that p principal components are derived but data reduction is achieved with first k PC's.

Do It Skills
do it. enjoy it.

PCA: GENERAL CONCEPT…

- From p original variables: $x_1, x_2, \ldots, x_p$, derive p new variables $y_1, y_2, \ldots y_p$, :

$$y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1p}x_p$$

$$y_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2p}x_p$$

…

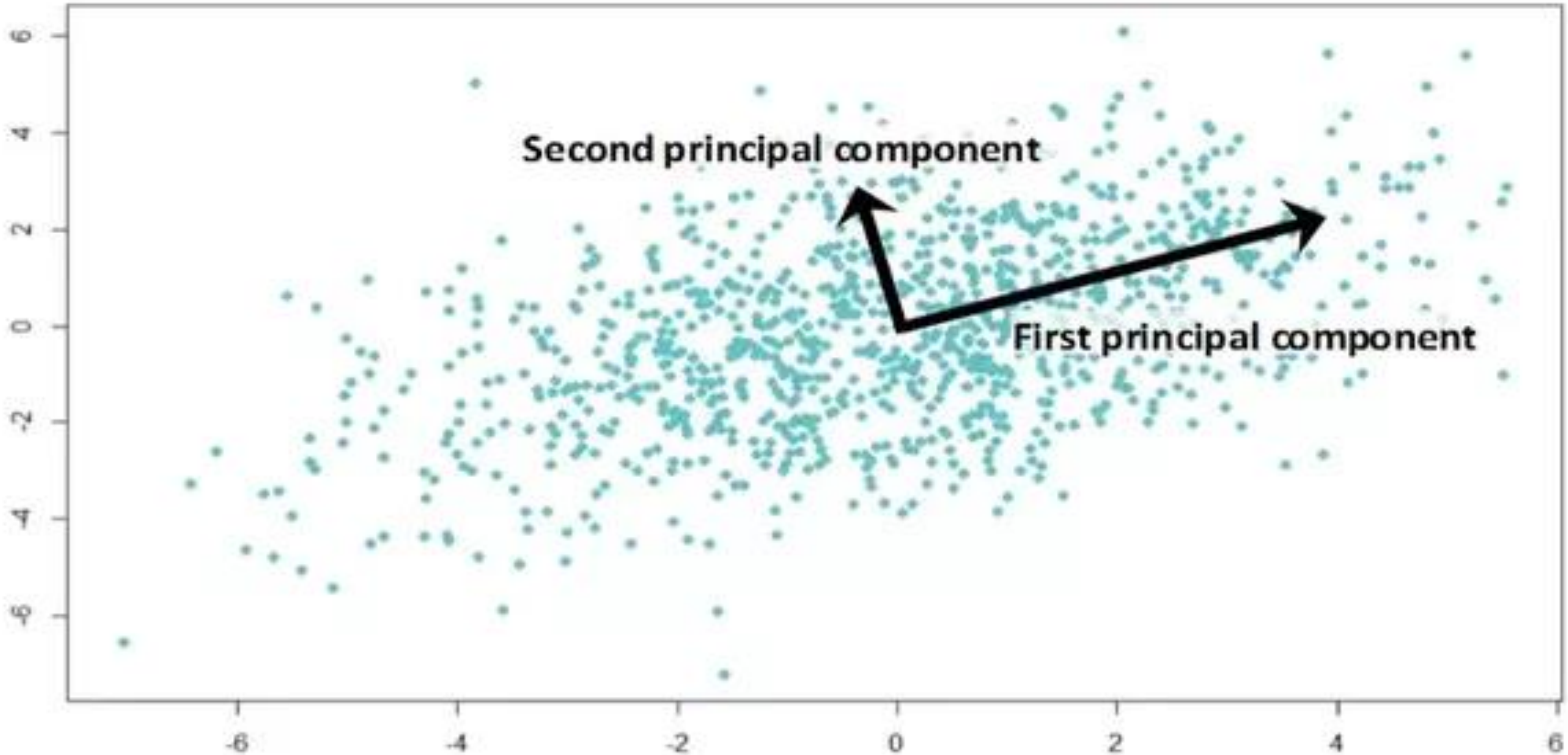
$$y_p = a_{p1}x_1 + a_{p2}x_2 + \ldots + a_{pp}x_p$$

There are y_k's Principal Components

such that:

- y_k 's are uncorrelated (orthogonal)
- y_1 explains as much as possible of original variance in data set
- y_2 explains as much as possible of remaining variance etc.

PCA PLOT


Second principal component
First principal component
6
4
2
0
-2
-4
-6
-6
-4
-2
0
2
4
6

PCA: CHOICE OF MATRIX FOR ANALYSIS

- Variance-Covariance matrix can be used if variables are measured on same units. Even then variables with high variances will dominate principal components.
- Correlation matrix is a better choice for PCA than variance-covariance matrix.
- Note that variance-covariance matrix of standardized variables is same as correlation matrix of original variables.

PRINCIPAL COMPONENTS: FORMAL DEFINITION

- First Principal Component: A linear combination a1'X which maximizes V(a'X) subject to a'a=1.
- Second Principal Component: A linear combination a2'X which maximizes V(a'X) subject to a'a=1 and Cov(a2'X and a1'X)=0
- Third Principal Component: A linear combination a3'X which maximizes V(a'X) subject to a'a=1, Cov(a3'X and a1'X)=0, Cov(a3'X and a2'X)=0

and so on..

MATRIX THEORY BEHIND PRINCIPAL COMPONENTS

Obtain correlation matrix R of ‘p ’analysis variables

Let

$(\lambda_1, e_1),(\lambda_2, e_2)......................(\lambda_p, e_p)$ be the eigenvalue-eigenvector pairs

where

$(\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq .... \geq \lambda_p \geq 0)$

Do It Skills
do it. enjoy it.

MATRIX THEORY BEHIND PRINCIPAL COMPONENTS..

- The ith Principal Component is given by

$y_i = e_i'x$

$= e_{1i}x_1 + e_{2i}x_2 + \ldots\ldots\ldots\ldots\ldots + e_{pi}x_p \qquad i=1,2,3.....p$

where ,

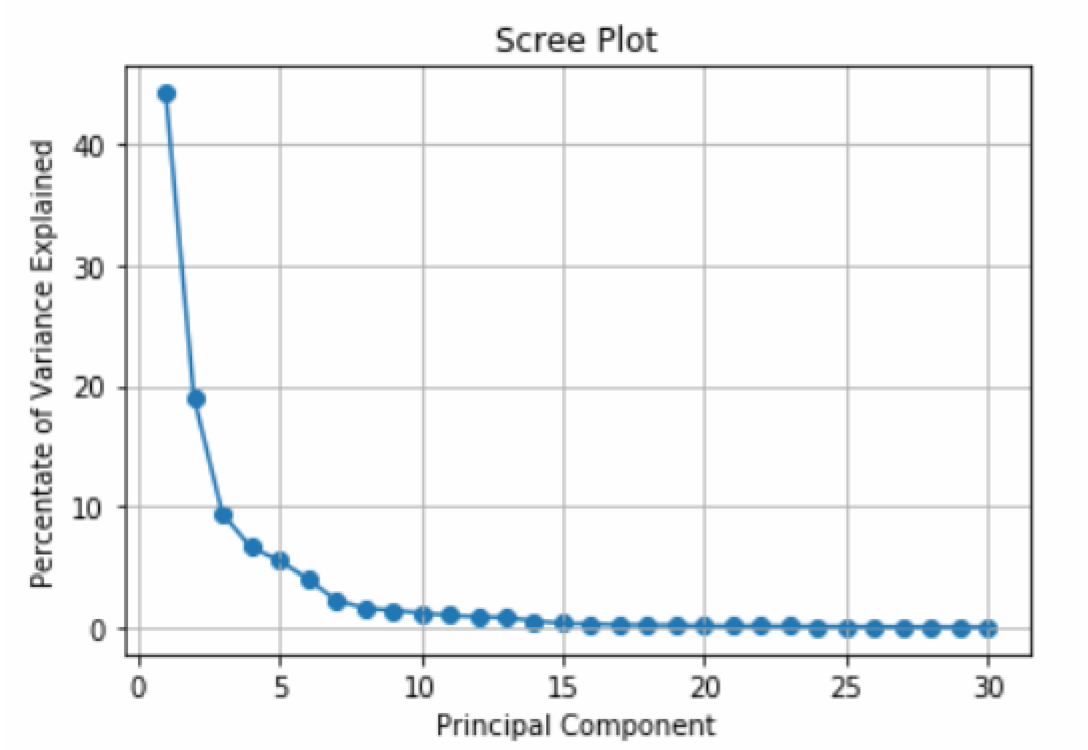
$V(y_i) = \lambda_i \qquad i=1,2,3.....p$

$COV(y_i, y_K) = 0 \qquad i \neq k$

HOW MANY PRINCIPAL COMPONENTS TO RETAIN?

- Note that p principal components are derived but data reduction is achieved with first k PC's.
- One can retain principal components associated with eigen value more than '1'. Variance of principal component is equal to associated eigen value. This is known as "Kaiser Criterion".
- Other method is to look at a scree plot and check for "elbow" to determine the correct number of PCs to use.

SCREE PLOT



Scree Plot
Percentate of Variance Explained
40
30
20
10
0
0
5
10
15
20
25
30
Principal Component

Do It Skills

do it. enjoy it.

THANK YOU

ARUNKG99@GMAIL.COM

WWW.DOITSKILLS.COM