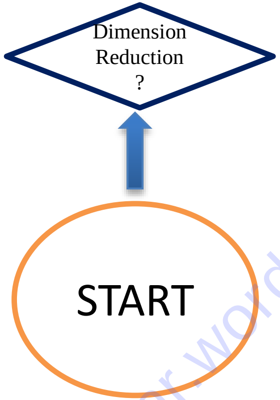
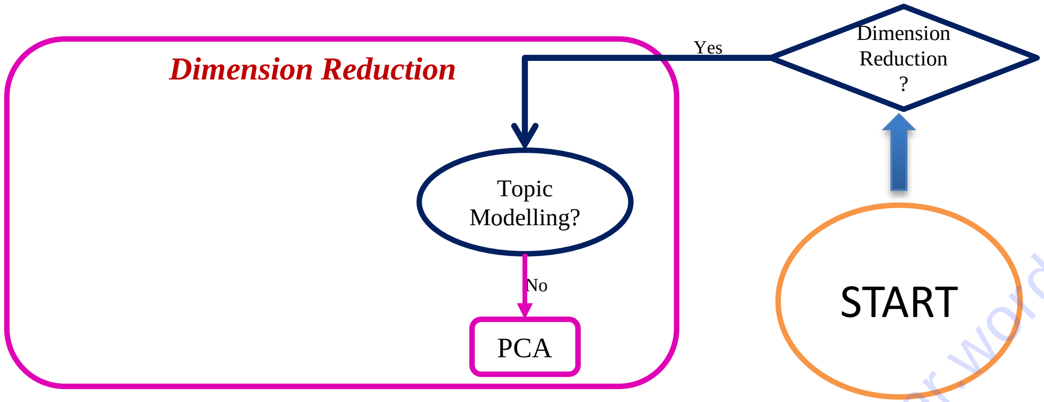
Machine Learning

Dr. Rajesh Kumar

PhD, PDF (NUS, Singapore)
SMIEEE, FIET (UK), FIETE, FIE (I),SMIACSIT, LMISTE, MIAENG
Professor, Department of Electrical Engineering
Professor, Centre of Energy and Environment
Malaviya National Institute of Technology, Jaipur, India, 302017
Tel: (91) 9549654481
http://drrajeshkumar.wordpress.com
rkumar.ee@mnit.ac.in, rkumar.ee@gmail.com


Dimension
Reduction
?
START

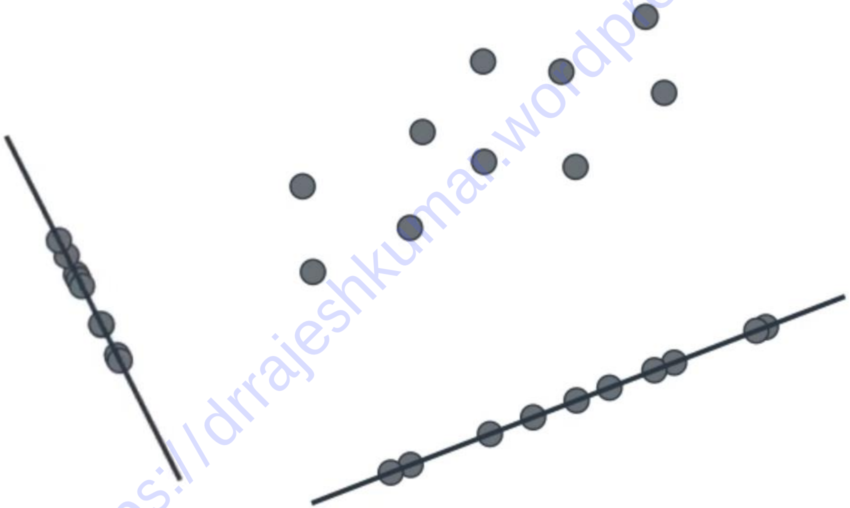

Dimension Reduction
Yes
Dimension Reduction ?
Topic Modelling?
No
PCA
START

Principal Component Analysis (PCA)

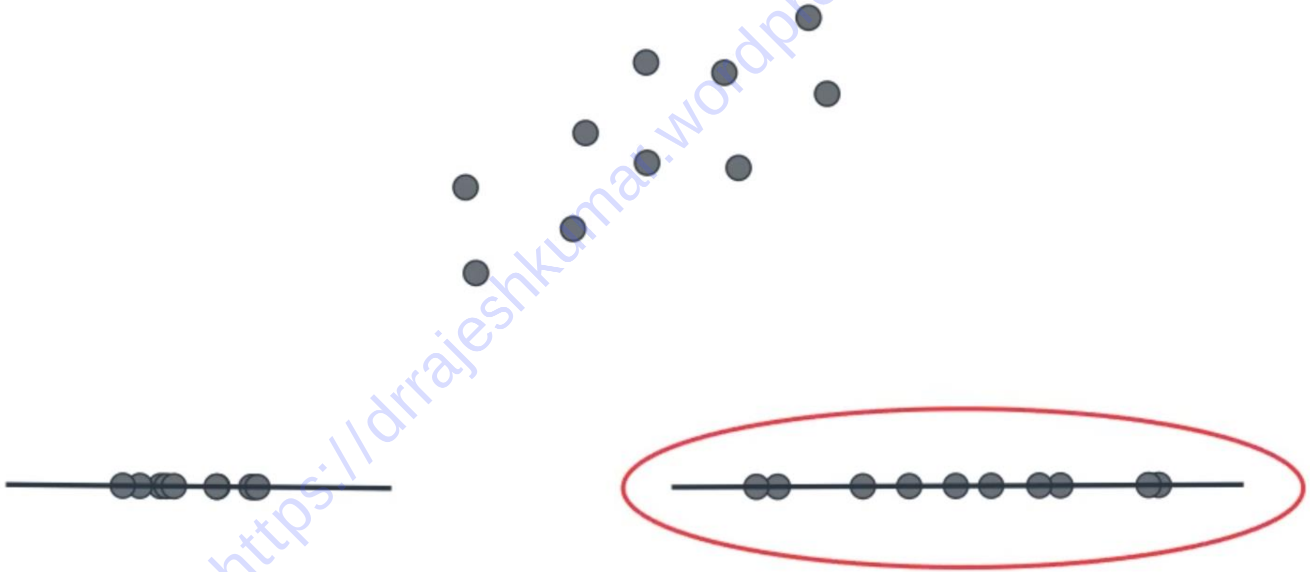
What is PCA?

- Unsupervised exploratory statistical technique used to:
  - **Simplify** and **reduce dimensionality** of the given dataset
  - **Visualize data** with high dimensionality
- Considers **maximized variance**
  - Maximum variance of a component ➔ a more significant factor
- Dimensions – different features that describe the data
  - Example: For 20 students, the number of hours studied and the marks obtained are provided.
    - Here number of hours and marks obtained are the dimensions

Dimensionality Reduction



Dimensionality Reduction



- Linear Transformation to determine a new coordinate system for the dataset
  - Greatest variance for any projection of the data set lies on the first axis → First Principal Component
  - 2nd greatest variance gives the Second Principal Component
- Dimensionality may be reduced by eliminating the principal components with least variance

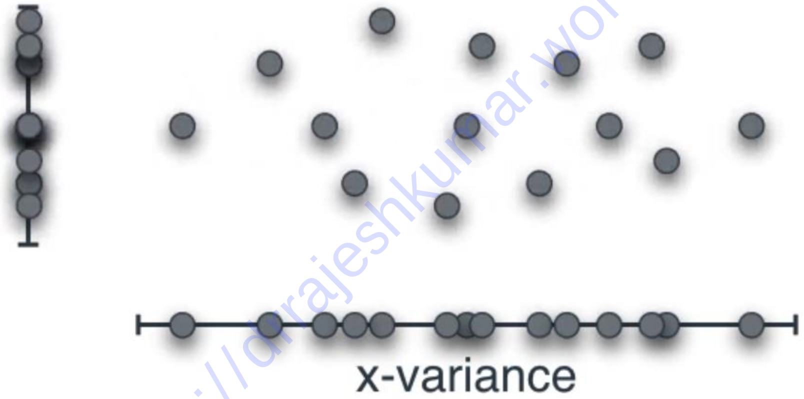
Variance and Covariance

- Measure how the data is distributed around the mean of that data.

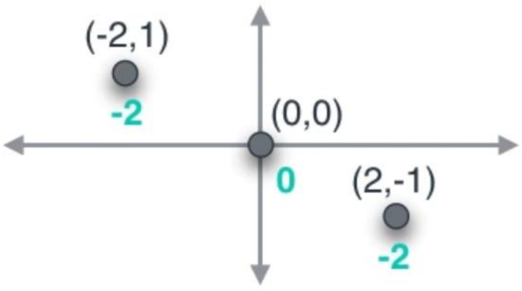
- **Variance** – Gives the deviation from the mean for data points in a single dimension.

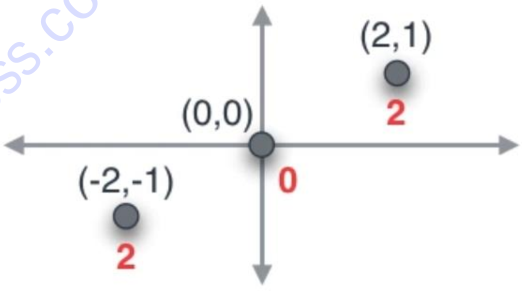
- **Covariance** – Measures the variation from the mean of each dimension with respect to the other.
    - Measures the existence of a relationship between 2 dimensions.
    - Square of the standard deviation

Variance?



Covariance





$$\text{covariance} = \frac{(-2) + 0 + (-2)}{3} = -4/3$$

$$\text{covariance} = \frac{2 + 0 + 2}{3} = 4/3$$

Covariance

- Covariance between two variables *x, y*

$$\boldsymbol{cov}(\boldsymbol{x},\boldsymbol{y}) = \frac{\sum_{i=1}^{n}(\overline{\boldsymbol{x}}_i - \boldsymbol{x})(\overline{\boldsymbol{y}}_i - \boldsymbol{y})}{\boldsymbol{n}-\boldsymbol{1}}$$

- *Covariance matrix*: represents the covariance between dimensions
  - Forms a symmetric *N X N* matrix for N dimensional data
  - Diagonal gives the variances of the dimensions

Covariance

- Covariance between two variables *x, y*

$$\boldsymbol{cov(x,y) = \frac{\sum_{i=1}^{n}(\bar{x}_i - x)(\bar{y}_i - y)}{n-1}}$$

- *Covariance matrix*: represents the covariance between dimensions
  - Forms a symmetric *N X N* matrix for N dimensional data
  - Diagonal gives the variances of the dimensions

$$\boldsymbol{C = \begin{bmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{bmatrix}}$$

Covariance matrix for 3 dimensions

Covariance

- Covariance between two variables *x, y*

$$cov(x, y) = \frac{\sum_{i=1}^{n}(\overline{x}_i - x)(\overline{y}_i - y)}{n-1}$$

$$C = \begin{bmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{bmatrix}$$

Covariance matrix for 3 dimensions

- *Covariance matrix*: represents the covariance between dimensions
  - Forms a symmetric *N X N* matrix for N dimensional data
  - Diagonal gives the variances of the dimensions

Covariance Evaluation

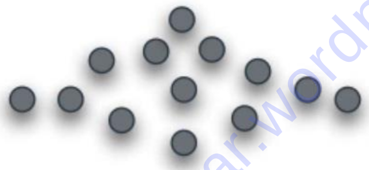
Covariance(x,y)	**Nature of Relationship**
= 0	*x*, *y* are **independent**
> 0	*x*, *y* move in **same direction**
< 0	*x*, *y* move in **opposing directions**

Covariance

$$\text{covariance} = \frac{-2 + 0 + 2 + 0 + 0 + 0 + 2 + 0 + -2}{9} = 0$$

Covariance

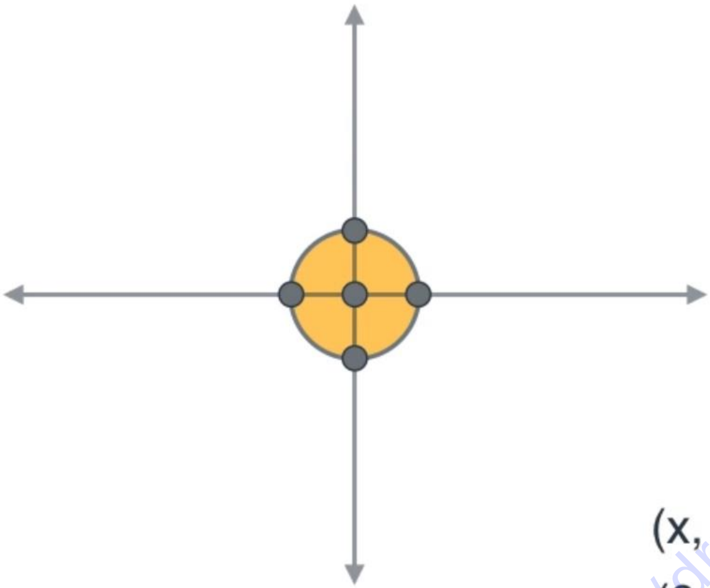
negative
covariance



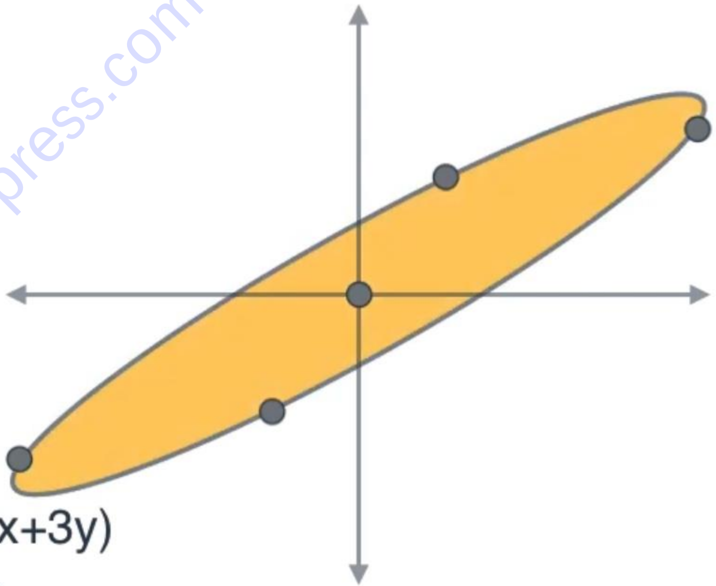
covariance zero
(or very small)

positive
covariance

Linear Transformations



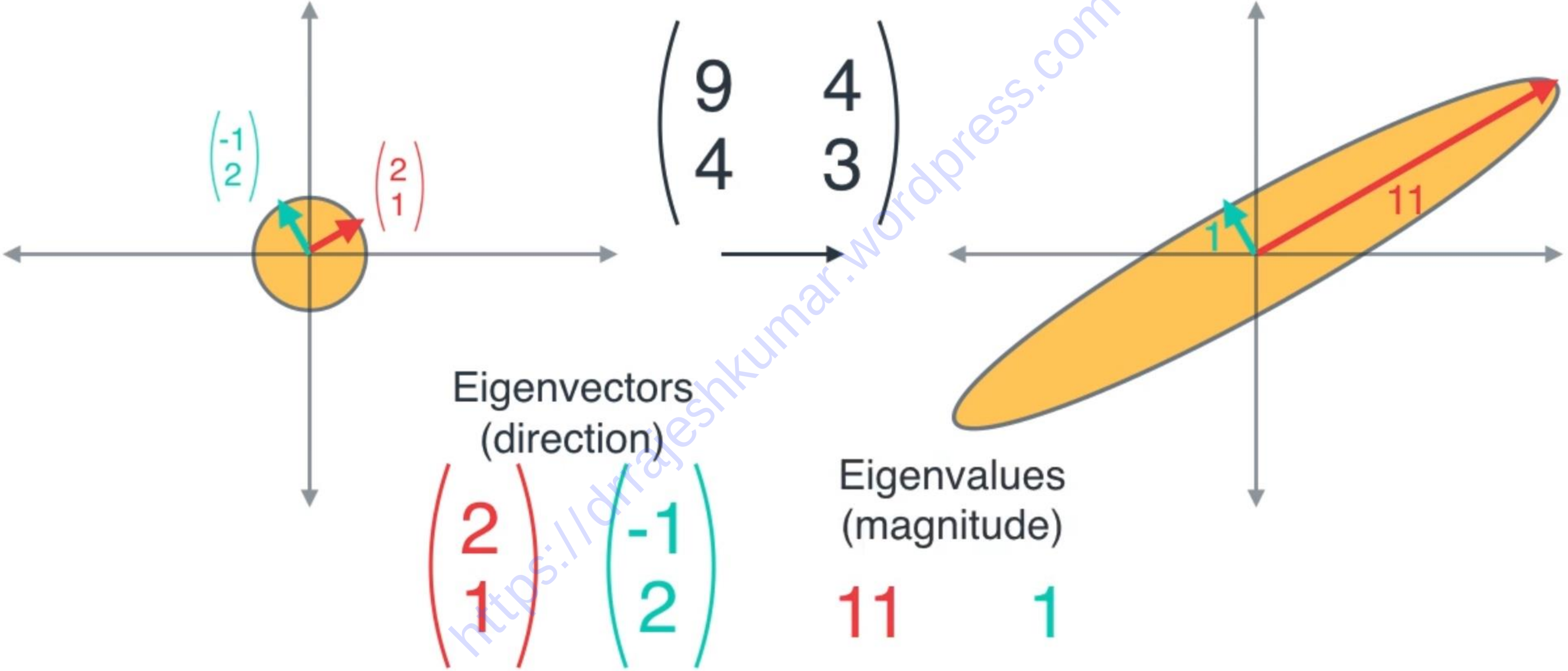
$$\begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix} \longrightarrow$$



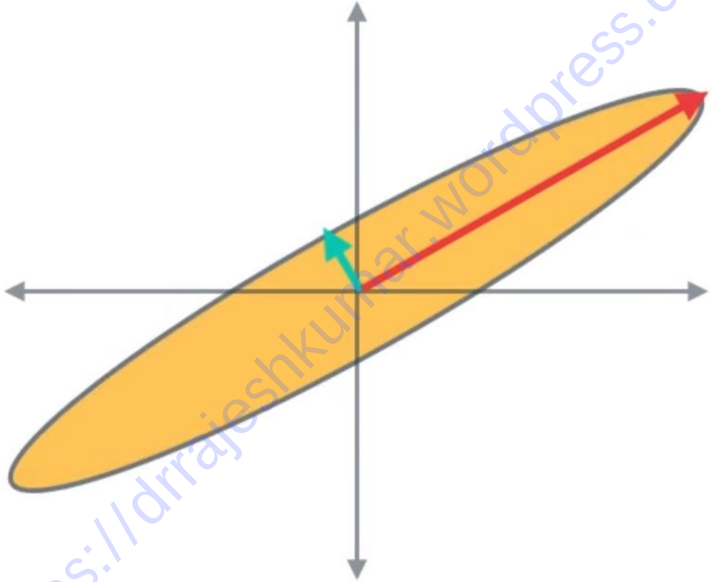
(x, y) ⟶ (9x+4y, 4x+3y)

(0,0)	(0,0)
(1,0)	(9,4)
(0,1)	(4,3)
(-1,0)	(-9,-4)
(0,-1)	(-4,-3)

Linear Transformations



Linear Transformations



Eigenvectors (direction)

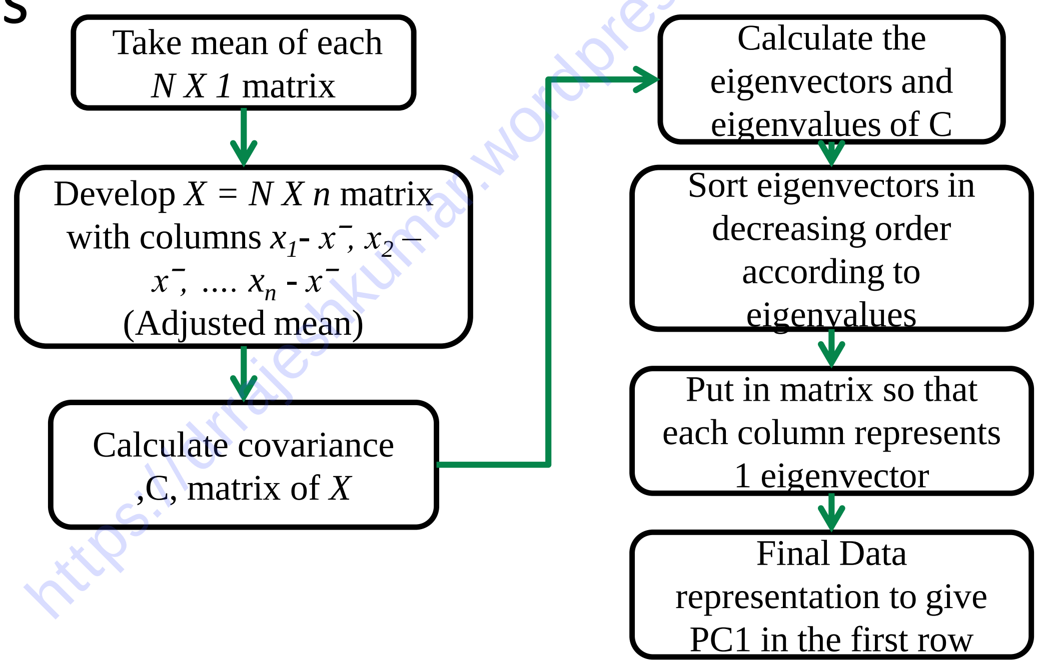
$$\begin{pmatrix}2\\1\end{pmatrix} \quad \begin{pmatrix}-1\\2\end{pmatrix}$$

Eigenvalues (magnitude)

11   1

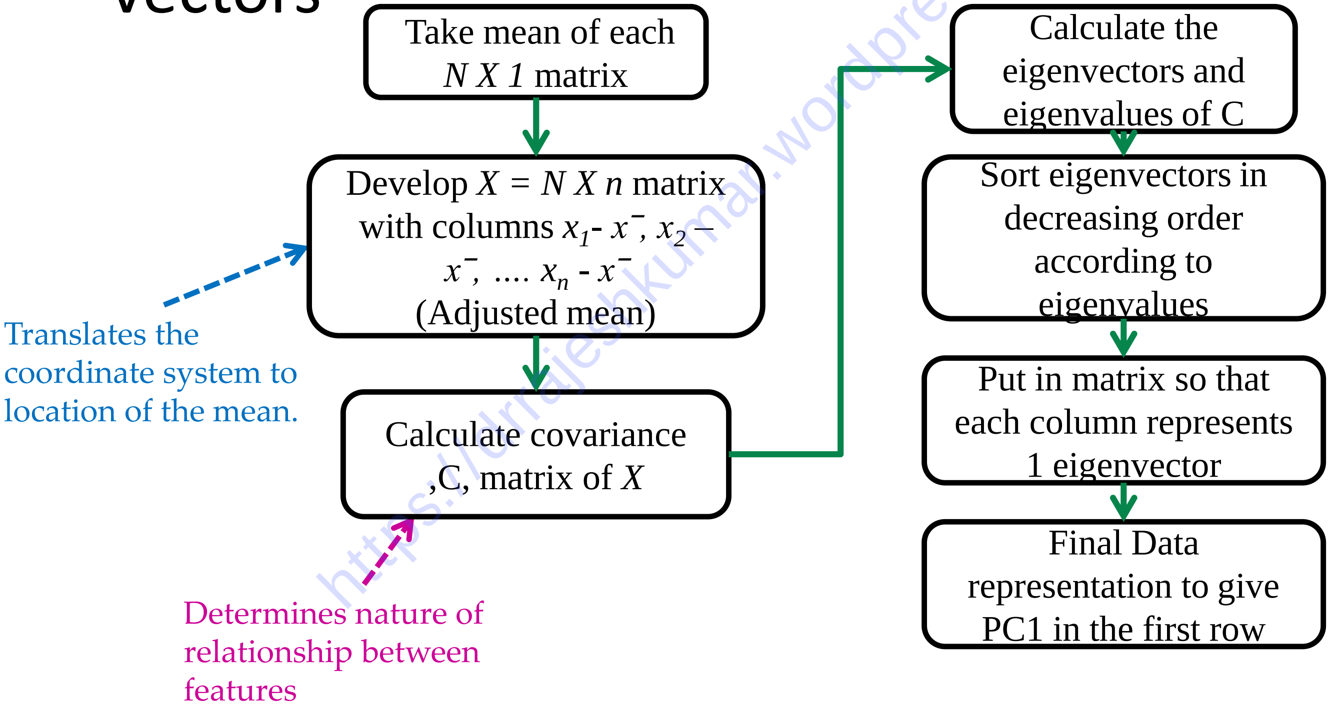
Steps for PCA - Flowchart

- Given $x_1, x_2, \ldots, x_n$ is a set of *n (N X 1)* vectors



Steps for PCA - Flowchart

- Given $x_1, x_2, \ldots, x_n$ is a set of *n (N X 1)* vectors



Steps for PCA - Flowchart

- Given x_1, x_2, ….,x_n is a set of *n (N X 1)* vectors

1. Take mean of each *N X 1* matrix
2. Develop $X = N \times n$ matrix with columns $x_1 - \bar{x}$, $x_2 - \bar{x}$, …. $x_n - \bar{x}$ (Adjusted mean)
3. Calculate covariance ,C, matrix of *X*
4. Calculate the eigenvectors and eigenvalues of C
5. Sort eigenvectors in decreasing order according to eigenvalues
6. Put in matrix so that each column represents 1 eigenvector
7. Final Data representation to give PC1 in the first row

Eigenvector with the highest eigenvalue gives the Principal Component .

Steps of PCA

- Given $x_1, x_2, ..., x_n$ is a set of *n (N X 1)* vectors

$$x_i = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{iN} \end{bmatrix}$$

- Calculate the average of each

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} x_{i1} \\ \vdots \\ x_{iN} \end{bmatrix}$$

- X is the *N X n* matrix with columns

$$X = [x_1 - \bar{x} \quad x_2 - \bar{x} \quad \dots \quad xn - \bar{x}]$$

This is the mean adjusted data

Steps for PCA

- Develop the covariance matrix by calculating the covariance matrix C from X

- Each term may be written as

$$x_j = \bar{x} + \sum_{i=1}^{i=n} g_{ji} eg_i$$

Steps for PCA

- Develop the covariance matrix by calculating the covariance matrix C from X

- Each term may be written as

$$x_j = \bar{x} + \sum_{i=1}^{i=n} g_{ji} eg_i$$

Eigenvector

- Calculate eigenvectors → $eg_1, eg_2, \ldots .eg_n$ will be *NX1* orthonormal vectors

Steps for PCA

- Develop the covariance matrix by calculating the covariance matrix C from X

- Each term may be written as

$$x_j = \bar{x} + \sum_{i=1}^{i=n} g_{ji} eg_i$$

Coordinates

- Calculate eigenvectors → $eg_1, eg_2, \ldots eg_n$ will be *NX1* orthonormal vectors
- Here, g_{ji} are the coordinates of x_j in the space

$$g_{ji} = (xj - \bar{x}).eg_i$$

Steps for PCA – Reduced Dimension Data Derivation

- Sort eigenvectors according to eigenvalue.
  - Matrix *E* gives the sorted eigenvalues with each column representing an eigenvector.
  - $E = [eg_1\ eg_2 \ldots\ eg_n]$

- Final Data Representation (Feature Vectors):
  - Determine **Row Feature Vector (RFV)** – matrix E transposed so that eigenvectors are <u>in rows</u>
    - Most significant eigenvector is at the top

  - Determine **Zero Mean Data (ZMD)** - Mean adjusted data matrix (X) transposed so that separate dimensions are in each row

$$\boldsymbol{Final\ Data = RFV \times ZMD}$$

Example

http://kybele.psych.cornell.edu/~edelman/Psych-465-Spring-2003/PCA-tutorial.pdf

Original Data

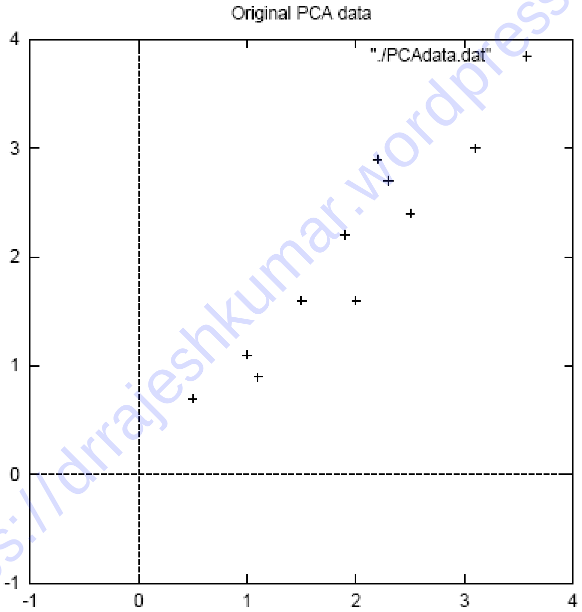
x	y
2.5	2.4
0.5	0.7
2.2	2.9
1.9	2.2
3.1	3
2.3	2.7
2	1.6
1	1.1
1.5	1.6
1.1	0.9

Adjusted Mean Data

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

Example

http://kybele.psych.cornell.edu/~edelman/Psych-465-Spring-2003/PCA-tutorial.pdf



Plot of the Data

Example

- Covariance matrix

$$C = \begin{bmatrix} .616555556 & .615444444 \\ .615444444 & .716555556 \end{bmatrix}$$

- Non-diagonal elements > 0
  - Variables increase/ decrease together.

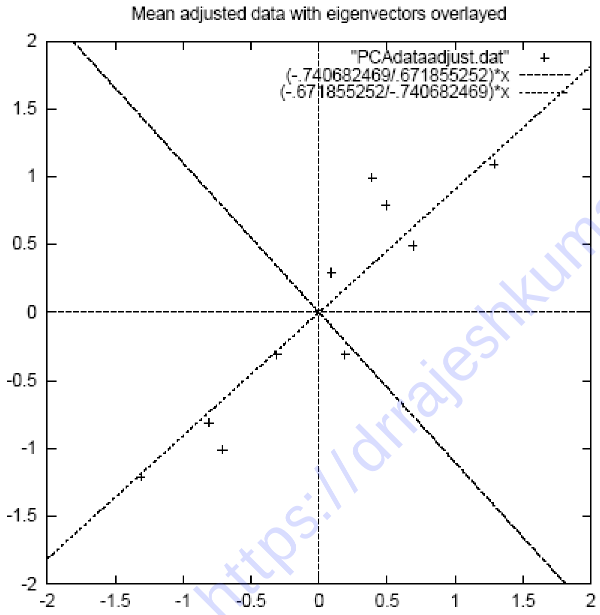
- Eigenvectors and eigenvalues of C

$$eigenvectors = \begin{bmatrix} .0490833989 \\ 1.28402771 \end{bmatrix}$$

$$eigenvectors = \begin{bmatrix} -.735178656 & -.677873399 \\ .677873399 & -.735178656 \end{bmatrix}$$

Example

http://kybele.psych.cornell.edu/~edelman/Psych-465-Spring-2003/PCA-tutorial.pdf



- One of the eigenvectors goes through the middle of the points, like drawing a line of best fit.
- The second eigenvector gives us the other, less important, pattern in the data.

Example

- **Principal Component → Eigenvector with the highest eigenvalue.**
- Order the eigenvectors by eigenvalue in decreasing order
  - This gives the significance of each eigenvector.
- The less significant vectors may be ignored.
  - May result in loss of information
  - Smaller the eigenvalues, lesser is the loss of information
- Final data set contains reduced dimensions!

Example – Feature Vector and Data

- Feature Vector with both eigenvectors ($[eg_2\ eg_1]$)

$$E = \begin{bmatrix} -.677873399 & -.735178656 \\ -.735178656 & .677873399 \end{bmatrix}$$

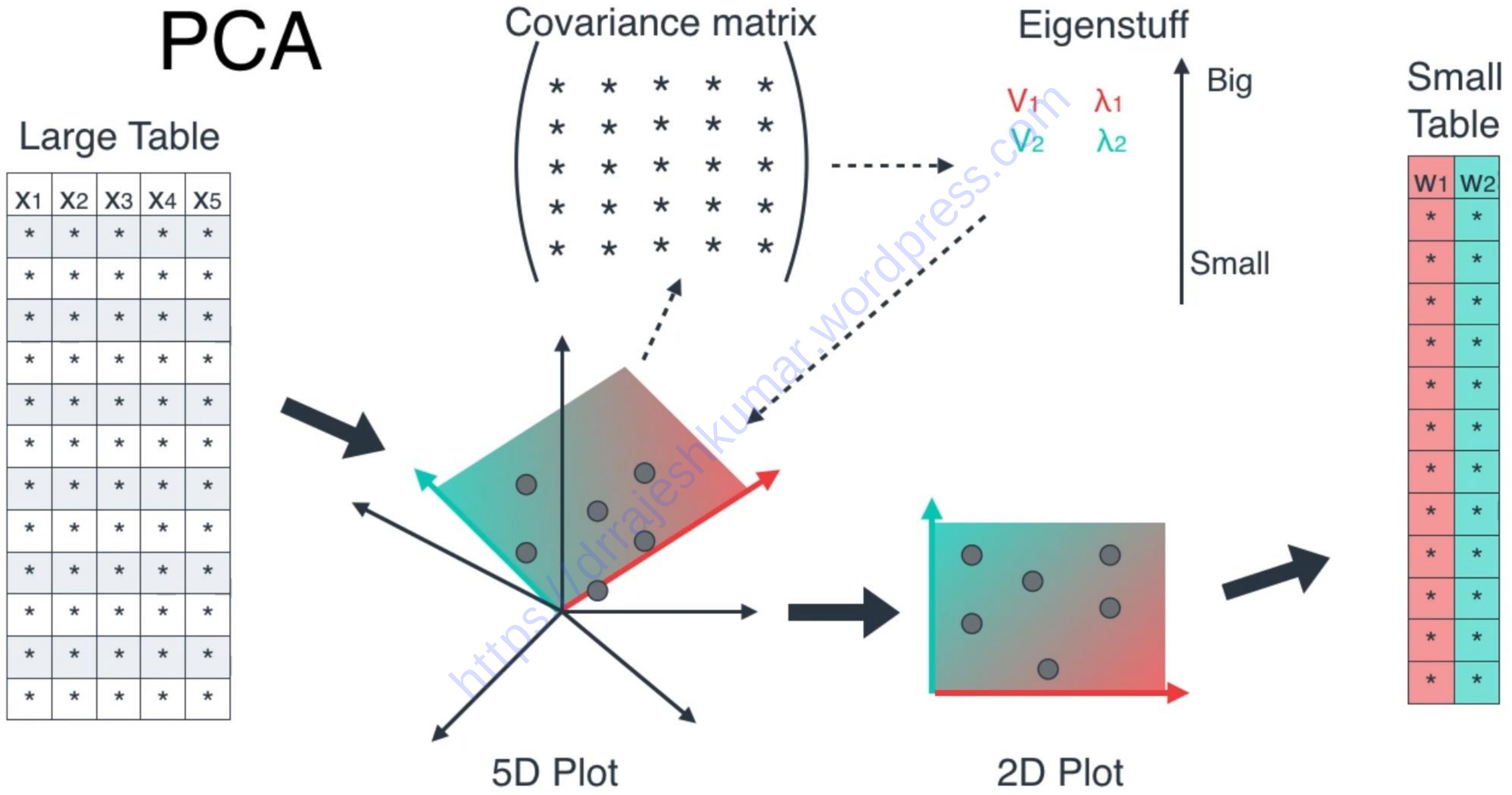
Can remove the less significant feature from matrix for reduced version.

- Data reconstruction with reduced features based on E and X!

$$\boldsymbol{Final\ Data = RFV\ \times ZMD}$$

Example – Final Data

x	y
-0.827970186	-0.175115307
1.77758033	0.142857227
-0.992197494	0.384374989
-0.274210416	0.130417207
-1.67580142	-0.209498461
-0.912949103	0.175282444
0.099109438	-0.349824698
1.14457216	0.046417258
0.438046137	0.01776463
1.22382056	-0.162675287


PCA
Large Table
x1 x2 x3 x4 x5
Covariance matrix
Eigenstuff
V1 λ1
V2 λ2
Big
Small
Small Table
w1 w2
5D Plot
2D Plot
https://drrajeshkumar.wordpress.com