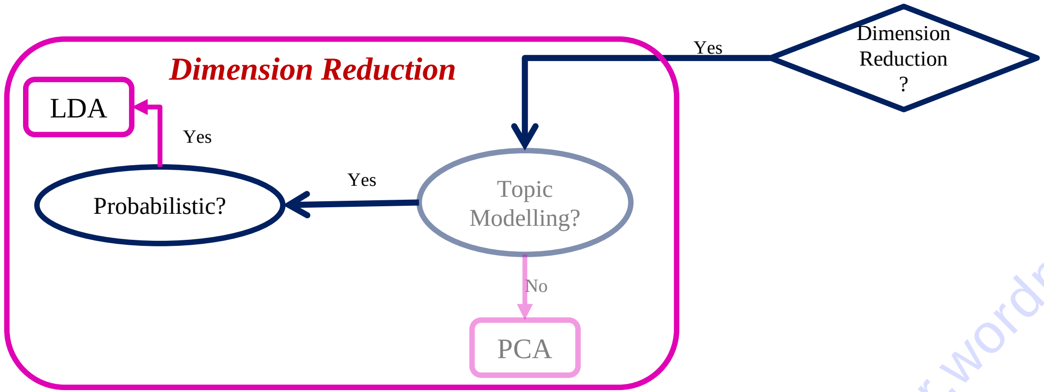
Machine Learning

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
Dimension Reduction ?
Yes
Topic Modelling?
Yes
Probabilistic?
Yes
LDA
No
PCA

Linear Discriminant Analysis (LDA)

Limitations of PCA

- Assumed that the 1st Principal Component also contains the most information
  - In some cases, **the highest variance may not be observed by the principal component.**

- Dependent on **linear assumptions**.

- Scale Variant
  - PCA does not affect the scale of the data
  - No data normalization in PCA
  - **Change in scale of one variable→ different results**

LDA

- Linear Discriminant Analysis
  - Supervised linear transformation
  - Used for maximizing the separation between multiple classes.
  - Used in feature extraction

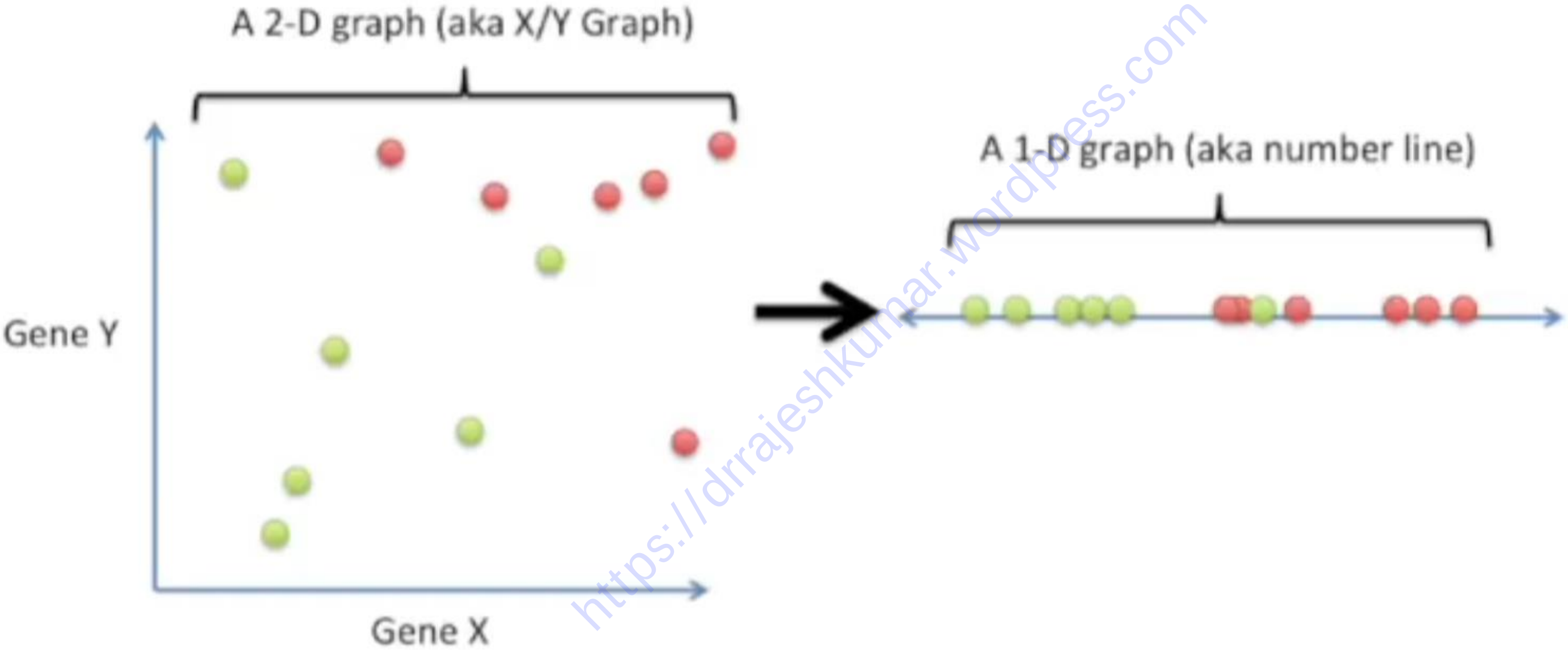
- LDA may be used as classifier and for dimensionality reduction.

LDA

- Assumptions in case LDA is used as classifier
  - Normally distributed data
  - Identical covariance matrices for each class

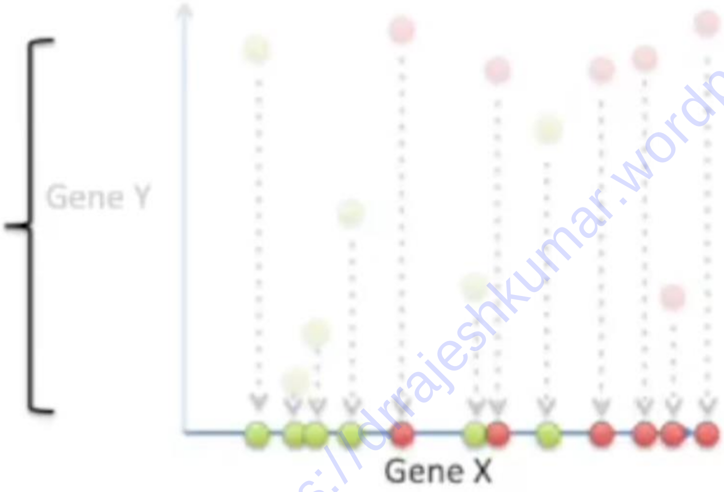
- Not applicable in case of dimensionality reduction

Reducing a 2-D graph to a 1-D graph



Reducing a 2-D graph to a 1-D graph

One bad option would be to ignore Gene Y

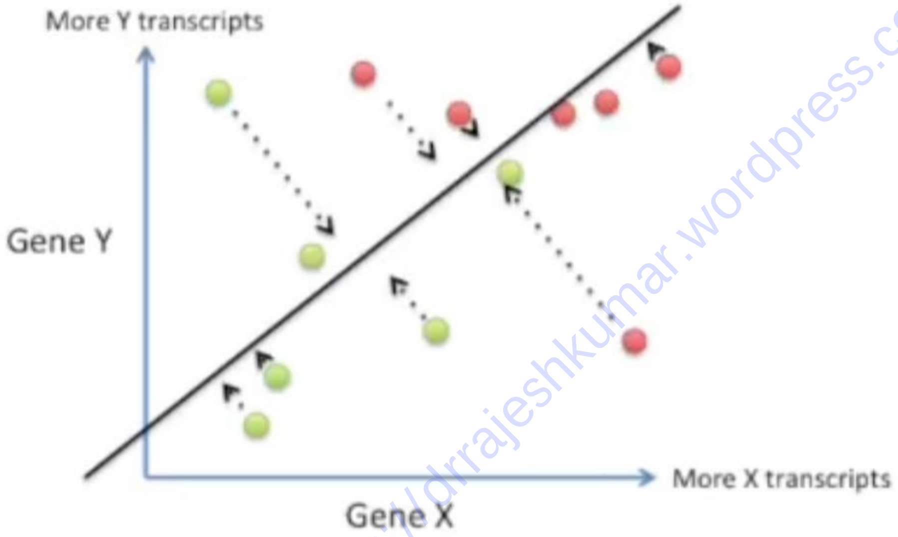


This way is bad because it ignores the useful information that Gene Y provides...

Projecting the genes onto the Y-axis (i.e. ignoring Gene X) isn't any better

Reducing a 2-D graph to a 1-D graph with LDA

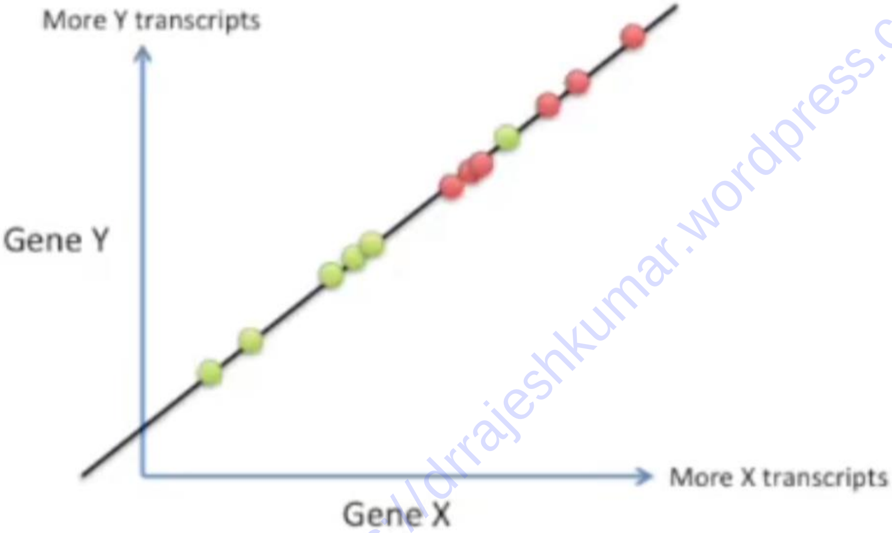
Reducing a 2-D graph to a 1-D graph with LDA



LDA uses both genes to create a new axis...

...and projects the data onto this new axis in a way to maximize the separation of the two categories.

Reducing a 2-D graph to a 1-D graph with LDA



LDA uses both genes to create a new axis...

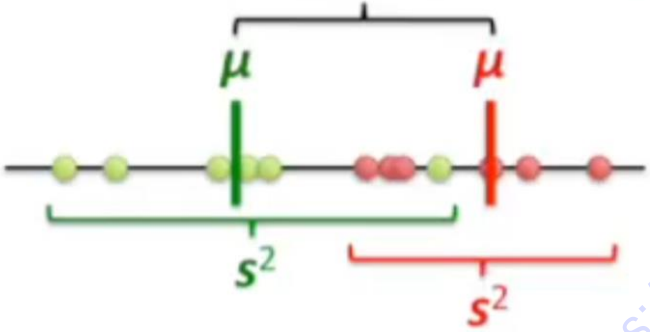
...and projects the data onto this new axis in a way to maximize the separation of the two categories.

How LDA creates a new axis...

The new axis is created according to two criteria (considered simultaneously):

1) Maximize the distance between means.

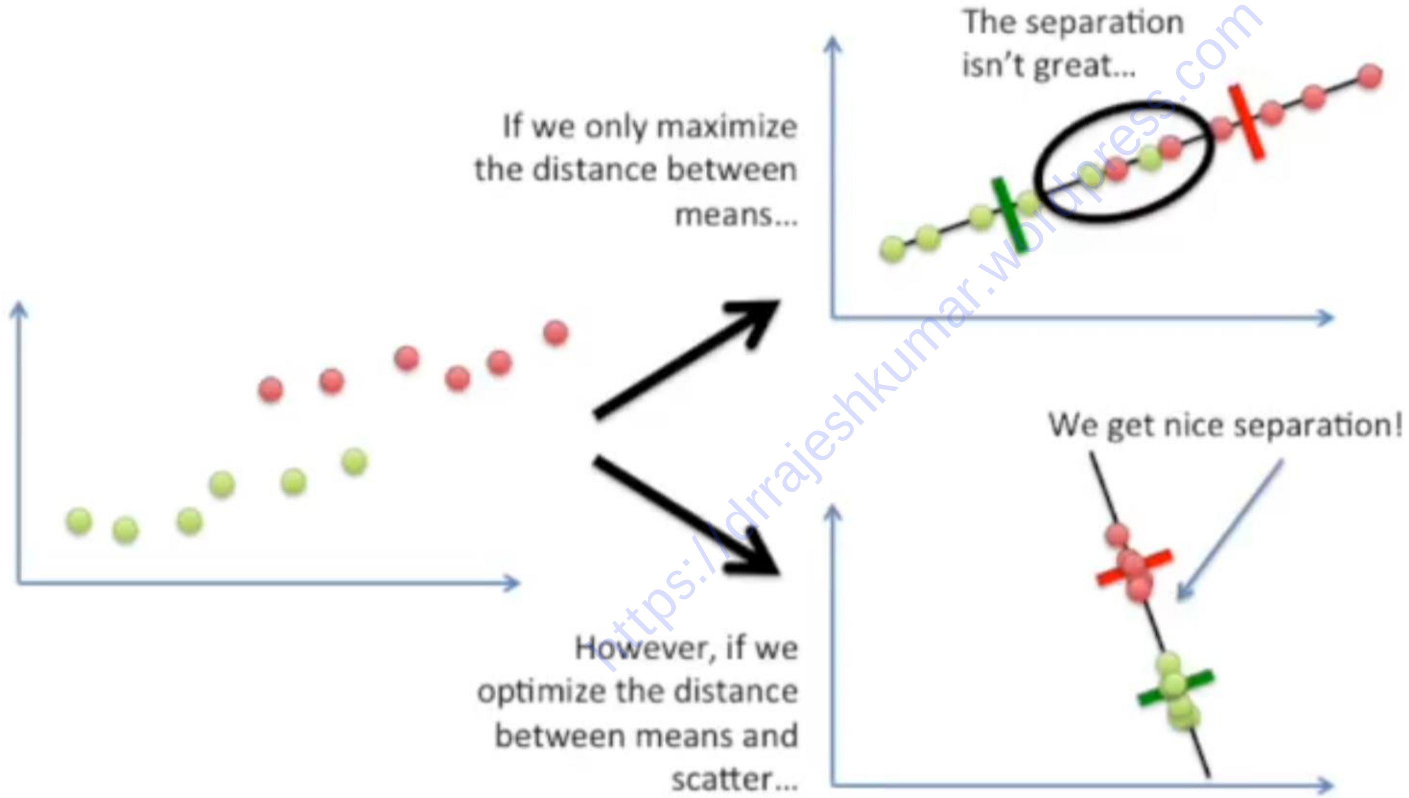
Let's call $(\mu - \mu)$ ***d*** for ***d***istance.



$$\frac{(\mu - \mu)^2}{s^2 + s^2} \rightarrow \frac{\text{Ideally large}}{\text{Ideally small}}$$

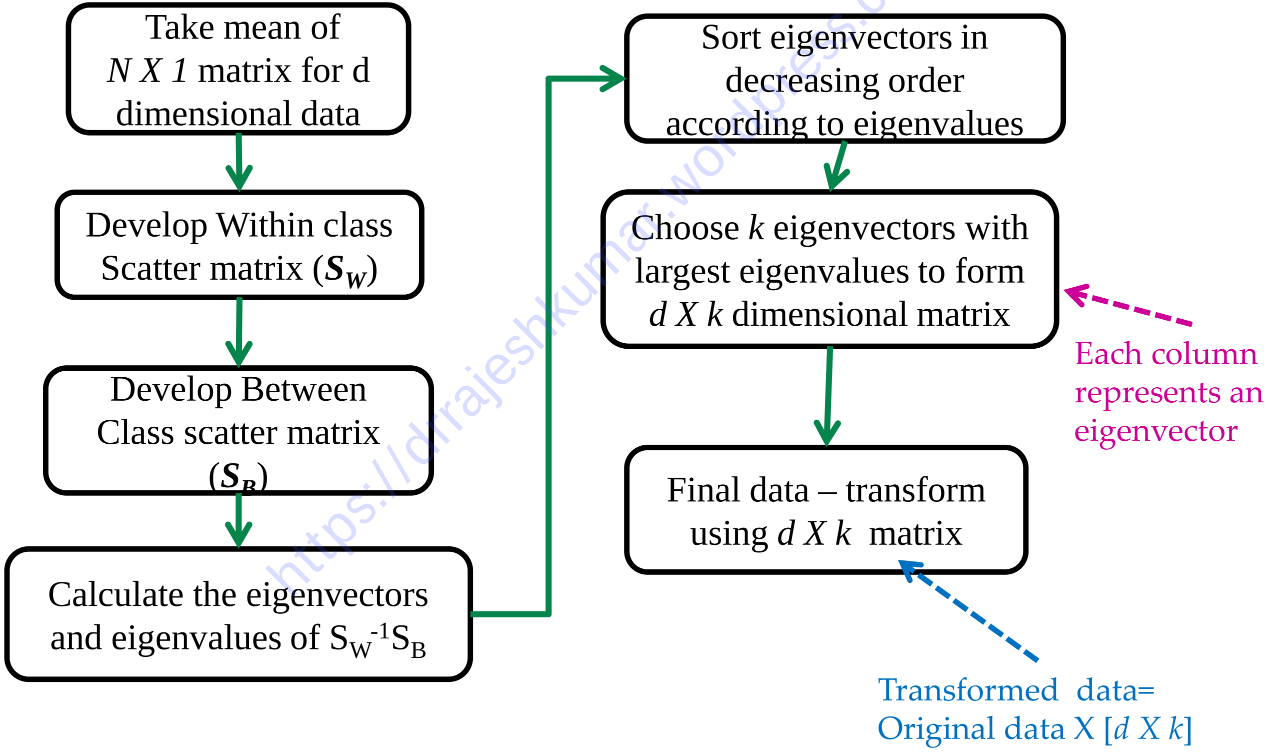
2) Minimize the variation (which LDA calls "scatter" and is represented by s^2) within each category.

An example showing why both distance and scatter are important.



Steps for LDA

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



Steps for LDA

- Consider an example with *n* classes and *N* features
- Calculate the mean vector m_n

$$\boldsymbol{m}_i = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_N \end{bmatrix}$$

Where *i = n = 1,2,3*

Steps for LDA

- Calculate the **within class scatter matrix S_W**

$$S_W = \sum_{i=1}^{c} S_i$$

where

$$S_i = \sum_{x \in Di}^{n} (\boldsymbol{x} - \boldsymbol{m}_i)(\boldsymbol{x} - \boldsymbol{m}_i)^T$$

Steps for LDA

- Calculate the **within class scatter matrix S_W**

$$S_W = \sum_{i=1}^{c} S_i$$

where

$$S_i = \sum_{x \in Di}^{n} (x - \boldsymbol{m_i})(x - \boldsymbol{m_i})^T$$

$$\boldsymbol{m}_i = \frac{1}{n_i} \sum_{x \in Di}^{n} \boldsymbol{x}_k$$

Steps for LDA

- Calculate the **within class scatter matrix S_W**

$$S_W = \sum_{i=1}^{c} S_i$$

where

$$S_i = \sum_{x \in Di}^{n} (\boldsymbol{x} - \boldsymbol{m}_i)(\boldsymbol{x} - \boldsymbol{m}_i)^T$$

- Calculate the **between class scatter matrix S_B**

$$S_B = \sum_{i=1}^{c} N_i(\boldsymbol{m}_i - \boldsymbol{m})(\boldsymbol{m}_i - \boldsymbol{m})^T$$

Steps for LDA

- Calculate the **within class scatter matrix S_W**

$$S_W = \sum_{i=1}^{c} S_i$$

where

$$S_i = \sum_{x \in Di}^{n} (\boldsymbol{x} - \boldsymbol{m}_i)(\boldsymbol{x} - \boldsymbol{m}_i)^T$$

- Calculate the **between class scatter matrix S_B**

$$S_B = \sum_{i=1}^{c} N_i(\boldsymbol{m}_i - \boldsymbol{m})(\boldsymbol{m}_i - \boldsymbol{m})^T$$

Overall mean

Steps for LDA

- Calculate the **within class scatter matrix S_W**

$$S_W = \sum_{i=1}^{c} S_i$$

where

$$S_i = \sum_{x \in Di}^{n} (\boldsymbol{x} - \boldsymbol{m}_i)(\boldsymbol{x} - \boldsymbol{m}_i)^T$$

- Calculate the **between class scatter matrix S_B**

$$S_B = \sum_{i=1}^{c} N_i(\boldsymbol{m}_i - \boldsymbol{m})(\boldsymbol{m}_i - \boldsymbol{m})^T$$

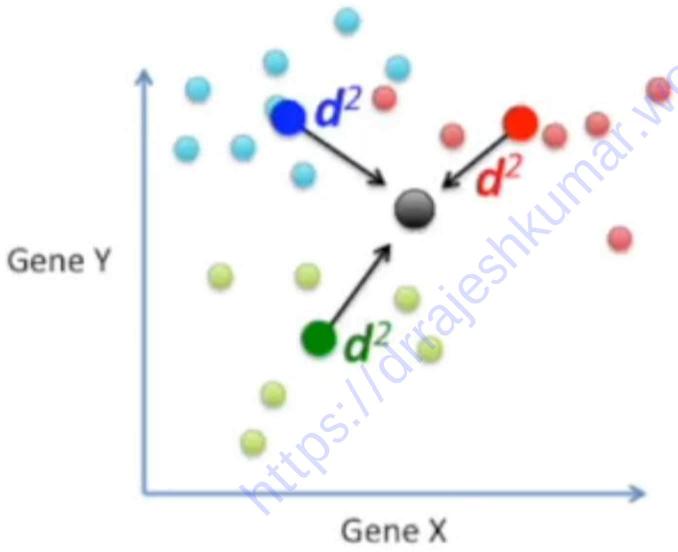
Sample size (N_i)

Sample mean ($\boldsymbol{m}_i$)

Steps for LDA

- Calculate eigenvectors and eigenvalues
  - Eigenvectors give the direction of the distortion of a linear transformation
  - Eigenvalues give the scaling factors for the eigenvectors
  - Eigenvalues → the axes for the new feature subspace
- Sort eigenvectors → choose eigenvectors with the largest eigenvalues
- Transform samples onto new subspace!

LDA for 3 categories



$$\frac{d^2 + d^2 + d^2}{s^2 + s^2 + s^2}$$

This is the same equation as before, but now there are terms for the **blue** category.

What is the difference between LDA & PCA?