# LOCALITY PRESERVING PROJECTION

Hari Haran .V

# Introduction

- *linear dimensionality reduction algorithm.*
- *Its an alternative to PCA – Principle component analysis.*
- *These are linear projective maps that arise by solving a variation problem that optimally preserves the neighborhood structure of the data set.*
- *LPP finds an embedding that preserves local information, and obtains a face space that best detects the essential manifold structure.*

# Dimensionality Reduction

- LINEAR METHODS:
  - PCA – Principle Component Analysis.
  - MDS – Multi Dimensional Scaling.
  - LDA- Linear discriminant Analysis.

- NON LINEAR METHODS:
  - ISOMAP-Isometric mapping methods.
  - LLE-Locally Linear Embedding.
  - Laplacian Eigen maps.

# Significance of LPP

- LPP is linear which is suitable for practical applications.
- The locality preserving quality is used in information retrieval applications.
- LPP is an approximation of non-linear methods that takes locality into account.
- It has higher accuracy when compared to PCA and LDA.

# ALGORITHM

## STEP 1 : **Constructing the adjacency graph**

Consider a graph G having k nodes and contains an edge between nodes i and j such that $X_i$ and $Y_j$ are close . There are 2 variations:

a) There is an edge between nodes i and j where $|| X_i - Y_j ||^2 < \varepsilon$ (neighbor hoods € $R^l$).

b) There is an edge between node if i is among n nearest neighbors of j Or j is among n nearest neighbors of i.

## STEP 2 : **Determination of Weights**

There are 2 variations :

a) Heat Kernel: : If there is an edge between 2 vertices i and j then ,

$$W_{ij} = e^{-\frac{\|\mathbf{X}_i - \mathbf{X}_j\|^2}{t}}$$

b) Simple Minded : If there is an edge connecting 2 vertices suppose i and j ,then $W_{ij} = 1$ .

## STEP 3 : **Eigen Maps**

Determination Eigen Vectors and Eigen values for the problem:

$$XLX^{T}w = \lambda XDX^{T}w$$

*Where* L is the Laplacian matrix, i.e. D – S, where S corresponds to the similarity values defined, and D is a column matrix which reflects how important a certain projection is. The more data points that surround a given point

Continue .....