

To implement the PCA algorithm and Fast Map Algorithm for Dimensionality Reduction.

PROJECT DESCRIPTION

- Programming language used : Python
- Data Structure : Lists
- File Name : PCA.py, FastMap.py
- Inputs: pca-data.txt, fastmap-data.txt, fastmap-wordlist.txt
- Output:
 - PCA – eigen vector, value for each dimension in 3-d space, 2-d Projection vector
 - Fast Map – Reduced coordinates of the objects in the projected hyperplane in 2d, Pivot points both before and after, plot for FastMap - wordlist

IMPLEMENTATION

- Modules created :
 - A. PCA
 - 1. ReadFile(): to save the input file into List of points for the given dimensions
 - 2. PCA(): to find the reduced coordinates based on the PCA algorithm and eigen values calculation.
 - B. Fast-Map
 - 1. ReadFile() : to save the input objects and their distances in the global variables ob1, ob2, dis , each of list type and further create M1[] matrix for distance of each object to the other.
 - 2. make2dList(rows, cols): to initialize a 2 dimensional list.
 - 3. DistantObjects(B,M): Chooses the most distant objects based on randomness, and linear heuristic algorithm approach, and gives Oa, Ob as the pivot points.
 - 4. D(i,M): Calculates distance between two objects, based on the distance matrix M using the formula:

$$x_i = \frac{d_{a,i}^2 + d_{a,b}^2 - d_{b,i}^2}{2d_{a,b}}$$

- 5. NewD(i, j, X): Calculates distance between object i and j for the new plane with projected distance X. Formula used:

$$(\mathcal{D}'(O_i', O_j'))^2 = (\mathcal{D}(O_i, O_j))^2 - (x_i - x_j)^2$$

- 6. FastMap(k,W) :
 - k – number of dimensions to reduce the data to, in our case k=2.
 - W is the distance matrix.

This module calculates the reduced dimensions based on the farthest point pivots in the given dimension, reduced dimension. The program runs recursively for 2 iterations (= number of reduced dimensions).

7. `f()`: calculates the new distance matrix from object 1 to 10 , to object 1 to 10 in the reduced plane.
8. `plotting()`: to plot the wordlist data on the reduced plane coordinates.

- Termination Condition :

- FastMap : When the dimension k is less than or equal to 1, it cannot be further reduced, module terminates.

- Result Interpretation:

- The PCA, FastMap output presents the reduced projection coordinates in a 2-d plane, for the points from the prior dimension.