DATA604 - HW11
Spring2024

Name: Swathi Baskaran
UID: 120428326

For my project, I selected the CIFAR-10 dataset, which has 60,000 data points. I shuffled the dataset and extracted a subset containing the first 20,000 data points.

The primary algorithm I opted for is Locally Linear Embedding (LLE). I executed LLE using the subset, 20,000 data points and measured the time it took for the algorithm to complete the computations.

Time taken to run the LLE algorithm: **67.14 seconds**

Cost of computation:

With the time calculated, the computational complexity of computing the LLE is broken down into 3 steps,

1. Nearest Neighbor Search: The time complexity depends on the implementation of the nearest neighbor search algorithm, which typically is around from **O[D log(K) N log(N)]**, where K nearest neighbors of N data points in D dimensions.

2. Weight Matrix Construction: The construction of the LLE weight matrix involves the solution of a K x K linear equation for each of the N local neighborhoods. The cost for this is approximately **O[D N K^3].**

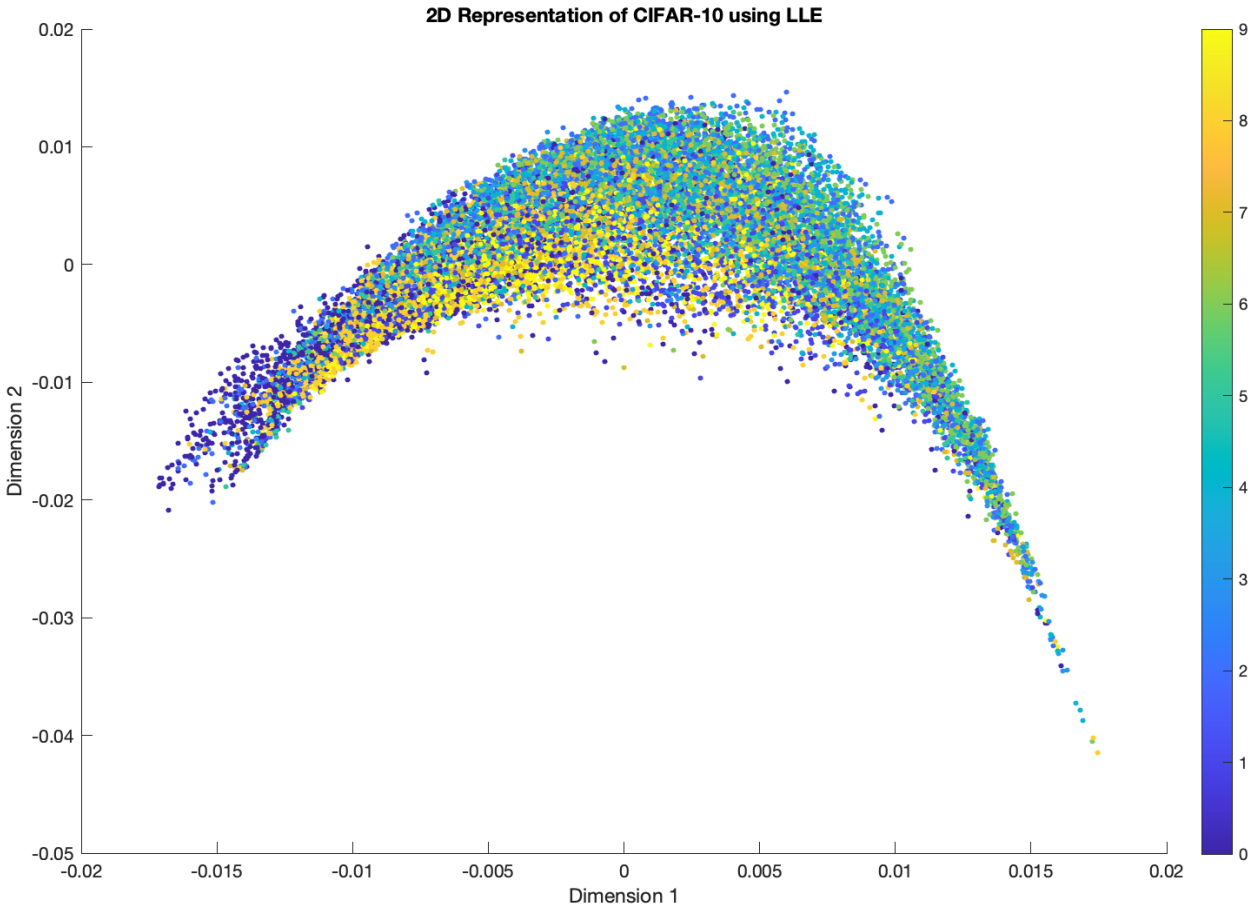
3. Eigenvalue Decomposition: This step involves computing the eigenvalues and eigenvectors of the weight matrix, typically with a time complexity of **O[d N^2]** depending on the output dimensionality of the data.

The overall cost of computation would be approximately:

$$O[D \log(K) N \log(N)] + O[D N K^3] + O[d N^2]$$

Where,
N = number of data points
K = number of nearest neighbors
D = input dimension
d = output dimension

Here is the 2-D image:



Here this time the algorithm took to run 20,000 data: 67.14 seconds

```
Command Window
  Finding nearest neighbors...
  Compute reconstruction weights...
  Compute embedding (solve eigenproblem)...
  Warning: Ignoring issym field in the options structure since the first input is not
  a function handle.
  > In eigs>checkIsSym (line 564)
  In eigs>checkInputs (line 381)
  In eigs (line 90)
  In lle (line 117)
  In compute_mapping (line 202)
  In HW11 (line 21)
  Warning: The first input matrix, shifted by sigma, is close to singular or badly
  scaled (RCOND =  6.680648e-19) and results may be inaccurate. Consider specifying a
  perturbed numeric sigma value to improve the condition of the matrix.
  > In eigs>WarnIfIllConditioned (line 1285)
  In eigs>AminusSigmaISolve (line 1238)
  In eigs>getOps (line 1150)
  In eigs (line 122)
  In lle (line 117)
  In compute_mapping (line 202)
  In HW11 (line 21)
  Time taken for LLE: 67.14 seconds
fx >>
```