**CSE 601 Project 1:**

**Dimensionality Reduction**

Team Member: Ninghui Jin

Instructor: Dr. A. Erdem Sariyuce

1. Nine scatter plots from three datasets and three algorithms. Label them properly by the dataset name and algorithm name in each plot.

* pca\_a.txt

A graph with red blue and purple dots

Description automatically generated

A graph of scatters

Description automatically generated with medium confidence

A graph with green and brown dots

Description automatically generated

* pca\_b.txt

A diagram of a scatter graph

Description automatically generated

A diagram of multiple colored dots

Description automatically generated

A diagram of multiple colored dots

Description automatically generated

* pca\_c.txt

A graph showing different colored dots

Description automatically generated

A graph of different colored dots

Description automatically generated

A graph of multiple scatters of cancer

Description automatically generated

1. Brief descriptions for the flow of your PCA implementation and discussion of the results obtained by different algorithms.

* Flow of PCA implementation.

1. Find the mean value for each column in the dataset.
2. Subtract the mean value of each column from the corresponding column, creating a mean-adjusted matrix.
3. Compute the covariance matrix for the mean-adjusted matrix.
4. Find the eigenvalues and eigenvectors of the covariance matrix.
5. Choose the top 2 eigenvalues and their corresponding eigenvectors.
6. The selected eigenvectors are the principal components.
7. Project the mean-adjusted data onto the principal components by taking the dot product of the transposed principal components and the transposed mean-adjusted data.

* Discussion of the results obtained by different algorithms.

The result obtained from PCA aligns with the direction of maximum variance, showcasing the projection onto the principal component. In contrast, the scatter plot created by SVD exhibits a wider spread. The shapes of the scatter points between them are somewhat similar, unlike t-SNE. Points generated by t-SNE tend to have larger values compared to other algorithms. Consequently, it becomes easier to discern the boundaries of different classes, as it preserves more of the relationships within the dataset.