**Lab 5: Unsupervised Learning Report**

**Introduction**

In Lab 5, we explored unsupervised learning techniques to uncover patterns within our dataset in the absence of labeled data. We applied a variety of clustering methods, including K-Means, Agglomerative Clustering, Spectral Clustering, MeanShift, Birch, and Affinity Propagation. Subsequently, we evaluated the quality of our clustering results using various metrics. To enhance our understanding of the outcomes, we visualized the clusters by reducing data dimensions to two through PCA and t-SNE. These visualizations provided a more intuitive representation of the clustering results.

**Clustering Results**

Our unsupervised learning experiment produced promising results, as demonstrated by the following evaluation metrics:

Silhouette Score: 0.4103

This score indicates that the clusters are well-defined and that data points are closer to members of their own cluster than to those in other clusters, reflecting the quality of our clustering.

Calinski-Harabasz Score: 6223.1544

A high score signifies that the clusters are distinct and well-separated, indicating the efficacy of our clustering methods in capturing underlying data patterns.

Davies-Bouldin Score: 0.9746

A low score suggests that the clusters are mutually exclusive and well-separated. Our clustering methods excelled in maintaining cluster separation.

Normalized Mutual Information (NMI): 0.0193

Although this value is relatively low, it suggests that there is a non-random association between the true and predicted cluster labels.

Adjusted Rand Index (ARI): 0.0008

This score, while close to zero, implies that the clustering methods performed slightly better than random chance.

Adjusted Mutual Information (AMI): 0.0080

AMI, like NMI, indicates a non-random association between the true and predicted clusters, albeit at a low level.

V-Measure: 0.0193

This balanced measure considers both homogeneity and completeness, providing further insight into the clustering performance.

Completeness Score: 0.1763

The completeness score measures how many data points that belong to the same true cluster are assigned to the same predicted cluster.

Homogeneity Score: 0.0102

The homogeneity score measures how many data points that belong to the same true cluster are assigned to the same predicted cluster.

**Visualization**

To complement our numerical evaluation, we harnessed dimensionality reduction techniques such as Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE). These techniques enabled us to visualize the clustering results in two dimensions, providing a more intuitive understanding of the data structure.

**PCA Visualization**

The PCA scatter plot showcased distinct clusters, revealing well-organized data points. Two clusters, in particular, were notably well-separated, affirming the efficacy of our chosen clustering algorithms.

**t-SNE Visualization**

It's worth noting that the t-SNE visualization, while used in this experiment, did not produce as strong results as expected. T-SNE is known to be more suitable for image datasets, and its performance may vary depending on the data characteristics. In our case, the t-SNE visualization did not exhibit as clear of a separation between clusters as the PCA visualization did.

In both visualizations, the predicted cluster assignments served as the third variable, making it easier to discern the different clusters. These visualizations offered strong evidence that our clustering methods successfully structured the data into meaningful groups, despite the absence of labeled information.

The visualizations from both PCA and t-SNE can be found in the "figures" folder, allowing for a detailed examination of the results.

**Conclusion**

The evaluation metrics and visualizations collectively attest to the effectiveness of the unsupervised learning techniques applied in Lab 5. Although some of the metrics yielded low scores, the clustering methods demonstrated the capacity to uncover patterns and group data points into distinct clusters. These insights hold potential for a multitude of real-world applications, including customer segmentation, anomaly detection, and recommendation systems, facilitating data-driven decision-making and personalized solutions.