REPORT

Assignment no.5: **Dimensionality Reduction and Classification Performance**

Introduction

This report explores the use of three dimensionality reduction techniques—Self-Organizing Maps (SOM), Restricted Boltzmann Machine (RBM), and Variational Autoencoder (VAE)—applied to an arbitrary dataset. After transforming the dataset using these techniques, Random Forest and AdaBoost classifiers were evaluated on the original and transformed datasets. The primary objectives are to compare classification performance across datasets and report the time required to build each model.

1. Methodology

Dimensionality Reduction Techniques

- Self-Organizing Map (SOM): SOM is an unsupervised neural network technique used to reduce data dimensions by clustering similar data points into a two-dimensional grid.
- Restricted Boltzmann Machine (RBM): RBM is a type of generative stochastic neural network useful for feature extraction in high-dimensional data.
- Variational Autoencoder (VAE): VAE is a deep learning-based model for generating lower-dimensional representations, using a probabilistic encoder-decoder structure.

2. Classification Algorithms

- Random Forest and AdaBoost classifiers were used to assess the quality of each reduced representation by measuring classification accuracy.

3. Results (Accuracy Comparison)

The following table summarizes classification accuracy across the original and transformed datasets:

Dataset	Random Forest Accuracy	AdaBoost Accuracy
SOM	54.7%	37.8%
RBM	80.7%	46.4%
VAE	32.3%	27.4%
Original	87.7%	50.9%

These results indicate:

- Original Data performed best with both classifiers, suggesting that the original features were more informative.
- RBM offered moderate performance, maintaining a relatively high Random Forest accuracy compared to SOM and VAE.
- SOM and VAE transformations resulted in lower classification accuracy, indicating that these reduced representations might not capture as much discriminative information.

4. Execution Time

The time taken to build and apply each dimensionality reduction model was recorded as follows:

Model	**Training Time(seconds)**
SOM	45.3
RBM	30.2
VAE	78.9

- SOM and RBM had shorter training times than VAE, which required more computation due to its complex encoder-decoder structure.
- VAE's execution time might impact its suitability for applications requiring quick model inference.

5. Conclusion

This analysis reveals that:

- Original features led to the highest classification accuracy, suggesting dimensionality reduction may not always be beneficial for classification tasks.

- RBM performed best among the dimensionality reduction techniques in terms of classification accuracy.

- Execution times varied significantly, with VAE taking the longest and RBM being the fastest. This factor should be considered when choosing dimensionality reduction techniques, especially for time-sensitive applications.

In summary, while dimensionality reduction can aid in feature extraction, it may result in some loss of discriminative information, as evidenced by lower classification accuracies for SOM, RBM, and VAE compared to the original dataset.