UNSUPERVISED MACHINE LEARNING

EXPERENTIAL LEARNING

NEIL CARDOZ

AIML B

23070126079

**1. Introduction**

Dimensionality reduction is an essential step in machine learning and data analysis, aimed at reducing the number of input variables while preserving significant information. Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Singular Value Decomposition (SVD), and t-Distributed Stochastic Neighbour Embedding (t-SNE) are widely used dimensionality reduction techniques. This analysis compares them based on their working process, adopted methods, computational complexity, and suitability for real-world applications.

**2. Comparison of Techniques**

| **Feature** | **PCA (Principal Component Analysis)** | **LDA (Linear Discriminant Analysis)** | **SVD (Singular Value Decomposition)** | **t-SNE (t-Distributed Stochastic Neighbor Embedding)** |
| --- | --- | --- | --- | --- |
| **Working Process** | Finds orthogonal axes (principal components) capturing the most variance in data. | Maximizes class separability by finding linear combinations of features. | Factorizes a matrix into singular vectors and singular values. | Maps high-dimensional data to a lower-dimensional space while preserving local similarity. |
| **Method Adopted** | Eigenvalue decomposition of the covariance matrix. | Eigenvalue decomposition of the scatter matrices. | Matrix factorization (A = UΣV^T). | Probabilistic, non-linear mapping using Kullback-Leibler divergence. |
| **Computational Complexity** | O(n^3) (depends on number of features) | O(n^3) (depends on number of classes) | O(mn^2) (for an m × n matrix) | O(n^2) per iteration |
| **Supervised/Unsupervised** | Unsupervised | Supervised | Unsupervised | Unsupervised |
| **Scalability** | Scales well with large datasets. | Works well with labelled data but limited to small class counts. | Efficient for large matrices. | Computationally expensive for large datasets. |
| **Loss of Information** | Minimal if choosing enough components. | May lose variance not related to class separability. | Dependent on rank approximation. | May distort global structure but preserves local similarity. |
| **Best Use Cases** | Feature extraction, noise reduction, visualization, and pre-processing in machine learning. | Classification problems (face recognition, handwriting recognition). | Recommender systems, natural language processing, image compression. | High-dimensional data visualization, clustering, and manifold learning. |

**3. Suitability for Real-World Applications**

1. **PCA** is ideal for exploratory data analysis, image compression, and feature extraction in predictive modelling.
2. **LDA** is preferred in supervised learning for classification tasks such as face recognition and medical diagnosis.
3. **SVD** is commonly used in recommender systems (latent semantic analysis) and noise reduction in signal processing.
4. **t-SNE** is widely used in data visualization to identify clusters in high-dimensional datasets such as genomics and NLP.

**4. Conclusion**

Each dimensionality reduction technique has its strengths and limitations. PCA and SVD are efficient for linear transformations and large-scale applications, whereas LDA excels in supervised classification tasks. t-SNE provides excellent visualization but is computationally expensive. Choosing the right technique depends on the problem's nature, dataset characteristics, and computational constraints.