ASSIGNMENT 4

1. Practical Component: Apply and Compare Dimensionality Reduction Techniques (75 Points)

Objective:
Apply multiple dimensionality reduction techniques to the MNIST dataset and analyze their effectiveness in representing handwritten digits in a 2D space.

Techniques to Implement:

- **t-Distributed Stochastic Neighbor Embedding (t-SNE)**
- **Principal Component Analysis (PCA)**
- **Auto-Encoder (Neural Network-Based)**

Tasks:

1. **Data Preparation:**
    - Load the MNIST dataset.
    - Normalize and preprocess the data as required for each technique.
2. **Dimensionality Reduction:**
    - Apply each of the four techniques to reduce the dimensionality of the dataset to 2D.
3. **Visualization:**
    - Generate 2D scatter plots for each technique, coloring the points based on their digit labels (0-9).

Analysis:

- **Representation Quality:**
    - Determine which digits are better represented by each technique.
    - Analyze the clustering and separation of different digit classes.
- **Technique Comparison:**
    - Compare the strengths and weaknesses of each method in the context of the MNIST dataset.
    - Discuss computational efficiency and scalability.
- **Interpretation:**
    - Provide insights into why certain digits are better or worse represented by specific techniques based on their underlying mechanisms.

Deliverables:

- **Jupyter Notebook (`.ipynb`):**
    - Well-documented code implementing all tasks.

    - Inline explanations and visualizations.
- **PDF Report:**
    - Introduction and methodology.
    - Visualizations of 2D representations.
    - Detailed analysis and comparison.
    - Conclusion summarizing key findings.

2. Literature Review: Summarize a Relevant Paper (25 Points)

Objective:
Provide a comprehensive summary of a seminal paper related to dimensionality reduction techniques.

Assigned Paper:

- **"Visualizing Data using t-SNE" by Laurens van der Maaten and Geoffrey Hinton** *(Journal of Machine Learning Research, 2008)*

Tasks:

1. **Summary:**
    - Outline the main objectives and contributions of the paper.
    - Explain the t-SNE algorithm and how it differs from other dimensionality reduction techniques.
2. **Critical Analysis:**
    - Discuss the advantages and limitations of t-SNE as presented in the paper.
    - Relate the concepts from the paper to the t-SNE technique used in the practical component.
3. **Applications:**
    - Highlight potential applications of t-SNE beyond the scope of the assignment.
    - Provide examples of how t-SNE has been utilized in other research or industries.

Deliverables:

- **Written Summary:**
    - Include the summary in the PDF report.
    - Ensure clarity, conciseness, and coherence.
    - Maximum length: 2 pages.

Additional Instructions:

- **Code-Walk Requirement:**
  - After submission, schedule a code-walk session with a TA within one week.
  - Contact any of the TAs to arrange a mutually convenient time.
  - **Note:** The assignment will **not** be graded without completing the code-walk.
- **Submission Guidelines:**
  - Submit both the `.ipynb` and `.pdf` files through the designated platform by the deadline.
  - Ensure all files are named appropriately (e.g., `Lastname_Firstname_Assignment.ipynb` and `Lastname_Firstname_Assignment.pdf`).
- **Academic Integrity:**
  - Ensure all work is original.
  - Properly cite any external resources or references used.