DS3030: Data Analytics Lab
Assignment 9

Date: Oct 06, 2025

Timing: 2:00 to 5:00 PM Max marks: 12

Instructions

- Submit one .ipynb file containing all answers named as [**student name**]_**assignment[number].ipynb**
- Write the **Part number** in **separate text blocks** before the answers.
- Write **justifications/comments** as required.
- Use appropriate libraries: **mlxtend** for association rules, **sklearn** for dimensionality reduction.

---

1 Association Rule Mining on Bakery Transactions

Dataset: Bakery Transaction Data (BreadBasket_DMS.csv)
Dataset Description:

- **Date** → Date of transaction
- **Time** → Time of transaction
- **Transaction** → Transaction ID (all items with same ID belong to one basket)
- **Item** → Product purchased

1. **Frequent Itemset Mining**

   Tasks:

   1. Remove rows where Item = 'NONE' and remove duplicate transactions.
   2. Transform the data into a **one-hot encoded binary matrix** (rows = transactions, columns = items).
   3. Generate **frequent itemsets** using the **Apriori algorithm** with `min_support = 0.02`.
   4. Sort the itemsets in **descending order of support**.

5. Visualize the **top 15 frequent itemsets** using a **horizontal bar chart**:
   - x-axis = support
   - y-axis = itemsets

(4)

2. **Association Rule Generation**

   Tasks:

   1. Generate **association rules** using confidence as the metric with a **minimum threshold of 0.3**.
   2. Sort the rules in **descending order of confidence**.
   3. Display the **top 10 rules** with columns: antecedents, consequents, support, confidence, lift.
   4. Visualize the rules using a **scatter plot**:
      - x-axis = support
      - y-axis = confidence
      - Include proper title and labels to assess rule strength

(4)

2 Dimensionality Reduction: PCA vs t-SNE

1. **Generate the following datasets:**

   1. **Swiss Roll** (use `sklearn.datasets.make_swiss_roll`)
   2. **Two Moons** with noise (use `sklearn.datasets.make_moons`)

   For EACH dataset, perform the following:

   1. **Visualize the original dataset**
      - 3D visualization for Swiss Roll
      - 2D visualization for Two Moons
   2. **Apply PCA** to project the data to 2 dimensions and visualize the result
   3. **Apply t-SNE** (with suitable hyperparameters) to reduce the data to 2 dimensions and visualize the result
   4. **Compare and interpret** both the results from PCA and t-SNE (Please refer next page for sample results)

(4)

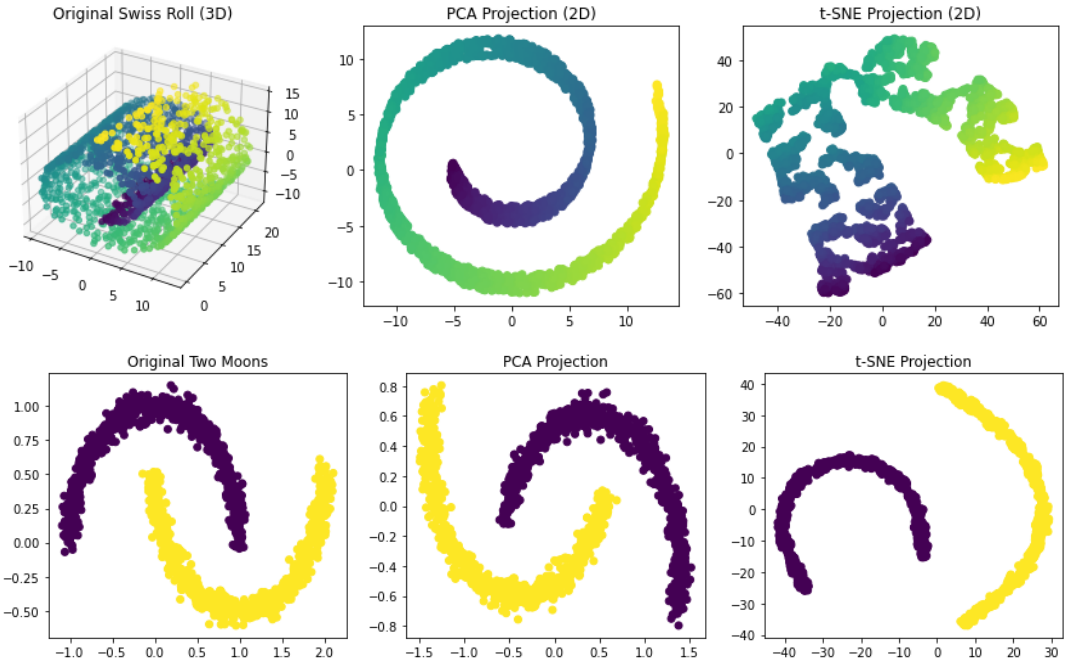

Figure 1: Results should be similar to the above: Swiss Roll (Top), 2-Moons (Bottom)