PES UNIVERSITY
CELEBRATING 50 YEARS

ML Lab Week 10 SVM Lab Instructions

1. Objective

The goal of this lab is to understand and implement Support Vector Machine (SVM) classifiers. You will train SVMs using three different kernels: **Linear, Radial Basis Function (RBF), and Polynomial**, on distinct datasets. You will then evaluate their performance using standard classification metrics and visualize their decision boundaries to see how they separate data.

---

2. Core Concepts

- **Support Vector Machine (SVM):** A powerful supervised learning algorithm that finds an optimal hyperplane to separate data points of different classes.
- **Kernel Trick:** A technique that allows SVMs to solve non-linear problems by transforming data into a higher-dimensional space.
  - **Linear Kernel:** Creates a straight-line decision boundary.
  - **RBF Kernel:** Creates a complex, non-linear boundary, like a circle or a wave.
  - **Polynomial Kernel:** Creates a curved, polynomial decision boundary.
- **Hard vs. Soft Margin:** The parameter `C` in SVMs controls the trade-off between maximizing the margin and minimizing the classification error. A large `C` leads to a hard margin (less tolerance for misclassification), while a small `C` leads to a soft margin (more tolerance).

---

3. Files Provided

- `boilerplate.ipynb` :A Jupyter Notebook file containing the skeleton code for the lab.

---

4. Lab Procedure

1. **Download and Setup:** Download the `boilerplate.ipynb` file and open it in your Jupyter Notebook environment.
2. **Review the Code:** Read through the entire notebook to understand the structure, the datasets being used, and the helper functions provided.

3. **Complete the Code:** Your main task is to fill in the sections marked with `TODO`. You will need to:
    - Initialize the `sklearn.svm.SVC` model with the correct kernel (`linear`, `rbf`, `poly`) or `C` value as specified in the comments.
    - Train the model using the `.fit()` method on the scaled training data.
    - Make predictions on the test set using the `.predict()` method.
    - Replace the placeholder `<PES1UG23CSXXX>` with your SRN wherever required.
4. **Execute the Notebook:** Run all the cells in the notebook from top to bottom. Ensure that all classification reports and decision boundary plots are generated and visible in the output.
5. **Analyze the Results:** Observe the performance metrics and the visualizations.
6. **Prepare Deliverables:** Once you have completed the notebook and answered the questions, prepare the two required files for submission as detailed in the next section.

---

5. Deliverables

You are required to submit **two separate files**. Please follow the naming conventions and content requirements carefully.

Deliverable 1: Completed Jupyter Notebook

This file demonstrates your coding work and the results you obtained.

- **File Naming Convention:** Rename your completed notebook to `SRN_SECTION.ipynb`.
    - *Example:* `PES1UG23CS123_A.ipynb`
- **Content Requirements:**
    1. **Code Completion:** All `TODO` sections in the notebook must be filled with the correct and functional code.
    2. **SRN Identification:** Your SRN must replace the `<PES1UG23CSXXX>` placeholder everywhere it appears (in the output of classification reports and in the titles of all plots).
    3. **Complete Output:** The notebook must be submitted with all cell outputs saved and visible. This includes the printed classification reports and all 8 generated plots. Do not clear the output before submitting.

Deliverable 2: Lab Report

This document contains your analysis and interpretation of the results.

- **File Format:** PDF Document.
- **File Naming Convention:** Name your report `SRN_SECTION_Report.pdf`.
    - *Example:* `PES1UG23CS123_A_Report.pdf`

- **Content Requirements:**
  1. **Cover Page:** Include your fullname, SRN, and section.
  2. **Analysis Questions:** Provide clear and concise answers to all 8 analysis questions from the notebook.The questions are divided into three sections:
     - **Moons Dataset Questions (2 questions):**
       1. Inferences aboutthe Linear Kernel's performance.
       2. Comparison between RBF and Polynomial kernel decision boundaries.
     - **Banknote Dataset Questions (2 questions):**
       1. Which kernel was most effective for this dataset?
       2. Why might the Polynomial kernel have underperformed here?
     - **Hard vs. Soft Margin Questions (4 questions):**
       1. Which margin (soft or hard) is wider?
       2. Why does the soft margin model allow "mistakes"?
       3. Which model is more likely to be overfitting and why?
       4. Which model would you trust more for new data and why?
  3. **Screenshots Provide clearly labeled screenshots for all the results generated by your notebook. You must include a total of** 14 screenshots, divided as follows:
  - **Training Results (6 Screenshots):** Capture the classification report output for each model.
    - **Moons Dataset (3 screenshots):**
      1. Classification Report for SVM with LINEAR Kernel with SRN
      2. Classification Report for SVM with RBF Kernel with SRN
      3. Classification Report for SVM with POLY Kernel with SRN
    - **Banknote Dataset (3 screenshots):**
    4. Classification Report for SVM with LINEAR Kernel
    5. Classification Report for SVM with RBF Kernel
    6. Classification Report for SVM with POLY Kernel
  - **Decision Boundary Visualizations (8 Screenshots):** Capture the plot for each model's decision boundary.
    - **Moons Dataset (3 plots):**
    7. Moons Dataset - SVM with LINEAR Kernel
    8. Moons Dataset - SVM with RBF Kernel
    9. Moons Dataset - SVM with POLY Kernel
    - **Banknote Dataset (3 plots):**
    10. Banknote Dataset - SVM with LINEAR Kernel
    11. Banknote Dataset - SVM with RBF Kernel
    12. Banknote Dataset - SVM with POLY Kernel
    - **Margin Analysis (2 plots):**

13. Soft Margin SVM (C=0.1)
14. Hard Margin SVM (C=100)

---

6. Submission

Submitthe two files (`SRN_SECTION.ipynb` and `SRN_SECTION_Report.pdf`) through the designated google form before the deadline. Please ensure to follow the naming conventions of the submission files.
