Week6 - Kernel SVM

DS3010: Introduction to Machine Learning Lab

Timing: 02:00 PM to 04:45 PM Max Marks: 5

Instructions

1. Submit one .ipynb file containing all answers. The name should be [student_name] _[rollno]_[lab].ipynb
2. Write the questions in separate text blocks before the answers.
3. Outputs for all sub-questions should be given and the code should be executable.
4. Write justifications for your choices where needed.
5. Ensure that all plots include clear labels and legends for better interpretation.
6. Use of generative AI tools (such as ChatGPT, Gemini, etc.) is strictly prohibited. Any submission found to contain AI-generated or plagiarized content will receive a score of zero, and disciplinary action.

You are given two datasets (data1.csv and data2.csv) with multiple numerical features and a categorical target label. Perform the following tasks on both the datasets:

1. **Data Preparation (1 mark)**
   - Load the dataset from CSV.
   - Normalize the feature values between 0 and 1.
   - Plot the dataset with different color for different classes.
   - Split the data into training (80%) and testing (20%).
2. **SVM Implementation (2 marks)**
   - Train SVM classifiers with the following kernels and parameters:
     i. Kernel - ['linear', 'rbf']
     ii. Vary C between 1e-5 to 1000 in multiples of 10.
     iii. Gamma - ['scale',' auto']
   - Report F1 scores for all models.
3. **Visualization (1 mark)**
   - Train the final model with the best hyperparameter based on F1 score and plot the decision boundaries.
   - Highlight misclassified points on the plots.
4. **Observation (1 mark)**

- Which kernel fits data1.csv and data2.csv the best respectively? Why?
- Does any kernel show signs of overfitting or underfitting in any of the given datasets? Why?
