TP 3 : Handwritten Digit Recognition with SVM

Problem:

Classify images of handwritten digits into their respective numerical values.

Data:

Use the "Digits" dataset available in scikit-learn.

Tasks:

1. Data Exploration and Pre-processing:
   - Load the Digits dataset from scikit-learn.
   - Explore the dataset to understand its structure and features.
   - Normalize and rescale the pixel values of the images.
2. Train an SVM Model:
   - Split the dataset into training and testing sets.
   - Implement SVM using scikit-learn with an appropriate kernel (e.g., linear or RBF) and using the one we saw in class (the SVM from scratch that we explained).
3. Model Evaluation:
   - Evaluate the SVM model's performance using metrics such as accuracy and F1 score.
   - Visualize the confusion matrix to understand the distribution of correct and incorrect classifications.
4. Visualize the Decision Boundary:
   - If applicable (for linear kernel), visualize the decision boundary learned by the SVM on a subset of the data.
5. Bonus: Classification of New Images:
   - Acquire or generate new images of handwritten digits.
   - Use the trained SVM model to predict the values of these new images.
   - Discuss potential challenges and considerations for deploying the model to new data.

Deliverables:

- Collab Notebook containing code for data exploration, pre-processing, model training, and evaluation.
- Visualizations illustrating the dataset features, confusion matrix, and any relevant insights.
- A report summarizing key findings, insights from the decision boundary analysis, and recommendations for deploying the model.

This TP focuses on digit recognition, a classic problem in image classification. The simplicity of the digits dataset allows for a clear understanding of SVM's application in image classification tasks.