

Assignment 2: Data classification using SVM

1. Objective

Support Vector Machine (SVM) is a classical supervised learning algorithm widely used for both classification and regression tasks. In this assignment, you will implement and analyze several variants of SVM on different datasets.

2. Tasks and Requirements

• Problem 1 (Linear classification basics):

Implement a linear-kernel SVM classifier on the Breast Cancer dataset.

- Clearly describe your data preprocessing and model training procedure (e.g., data preprocessing, train/test split, hyperparameter settings).
- Briefly interpret the results.

• Problem 2 (Comparison of different kernels)

Train SVM models using the linear kernel, polynomial kernel, and Gaussian RBF kernel on the Breast Cancer dataset.

- Summarize how different kernels and their parameter choices affect model accuracy.
- Is the position and orientation of the decision boundary reasonable?

• Problem 3 (Solving a more complex problem)

Conduct a SVM experiment on the MNIST dataset using an appropriate strategy. You may incorporate additional algorithms or techniques to improve performance (e.g., PCA, kernel optimization, or feature scaling). Based on the experiments, which two categories do you think have a blurry decision boundary?

3. Report and Submission

You should submit:

- all codes;
- your report (a separate PDF file).

Scikit tools:

[Support Vector Machines — scikit-learn 1.7.2 documentation](#)

Dataset Detail:

[Breast Cancer Wisconsin \(Diagnostic\) - UCI Machine Learning Repository](#)

[ylecun/mnist Datasets at Hugging Face](#)