

Theoretical Foundations of Machine Learning /2024

Course Project

You have to make Teams of **maximum 5 students**.

Datasets: MNIST, CIFAR-10, or Any other real Classification dataset of your choice (Other than Iris dataset)

Project Phases:

Phase 1:

- Select a dataset.
- Splitting: Prepare the train-validation and test portions.
- Data Preprocessing: Apply any preprocessing for features that you find suitable for the data.
- **Deadline: 2/12/2024**

Phase 2:

- Implement one case of SVM (Separable “Hard margin” or Non-separable “Soft margin” → depending on the dataset selection) from scratch as explained in the lecture “Not built-in”
Train and classify for SVM, you have to find optimal alpha, calculate W & b then formulate the classifier function. Then, draw the decision boundary if possible.
- Compute evaluation metrics [Accuracy, F1 score, recall, and precision]
- **Deadline: 16/12/2024**

Phase 3:

- Apply 3 Built-in models (SVM & any other 2 models)
- Compute evaluation metrics for each model [Accuracy, F1 score, recall, and precision]
- It is expected that the selected models should be experimented with different hyper-parameters.
- Compare between the four models [Implemented SVM, Built-in SVM, and the other 2 models] in a printed report and decide which one is better for your dataset
- **Deadline: 23/12/2024**

1- **MNIST**

It is a dataset of handwritten numbers from 0 to 9. MNIST has a training set of 60,000 examples, and a test set of 10,000 examples. It can be downloaded from:
<http://yann.lecun.com/exdb/mnist/>

2- **CIFAR-10**

CIFAR-10 is a dataset of 60000 color images downsized to 32*32 categorized in 10 classes. Each class contains 6000 samples. The dataset is divided into 50000 images for training and 10000 images for testing. It can be downloaded from:
<https://www.cs.toronto.edu/~kriz/cifar.html>