

# 3945 – Advanced Machine Learning, Spring 2022

## Home Assignment 1 – Unsupervised Learning

Lecturers: Dr. Leon Anavy, Mr. Alon Oring

---

In this assignment, you will explore unsupervised learning techniques for clustering and dimensionality reduction, and apply them to the MNIST database of handwritten digits. Your task is to explore the use of clustering methods and dimensionality reduction techniques as steps preceding to classification.

You will submit a single ipython notebook (.ipynb) file that includes all the code and outputs, as well as a brief report that explains your approach and findings. You are encouraged to use methods and techniques covered in class but are also welcome to use any other relevant method you find suitable for the task.

### Part 0: Load the data

- 1) Load the MNIST dataset.
- 2) Explore the data briefly

### Part 1: Clustering

- 3) For at least three different clustering methods of your choice perform:
  - a) Cluster the data using the raw pixel level values.
  - b) Evaluate the clustering result using standard evaluation metrics
  - c) Visualize the clusters by displaying a few images from each cluster.
  - d) Build a simple classifier based on the clustering results. Specifically, assign a label to each cluster (e.g., the most common digit in the cluster) and use this label to predict the digit of a test image based on its nearest cluster.

**Note:** You may suggest other methods to construct a classifier over the clustering results.

  - e) Evaluate the performance of the classifier.
  - f) Explore different parameters for the clustering and evaluate their effect on the results.

### Part 2: Dimensionality Reduction

- 4) For at least three different dimensionality reduction methods of your choice perform:
  - a) Reduce the dimensionality of the data and visualize the first few features in the new space.
  - b) Train at least one classifier on the data in the new space.
  - c) Evaluate the performance of the classifiers
  - d) Explore different parameters for the dimensionality reduction and evaluate their effect on the results.

### Part 4: Classification of the raw data

- 5) For at least one classification algorithm of your choice:
  - a) Train a classifier on the raw pixel level values.
  - b) Explore different parameters for the classification model
  - c) Evaluate the results of the model

### Part 3: Summary

- 6) Summarize the results obtained in parts 1-3 compare the results of different methods and suggest possible explanations.

## Submission Guidelines

- Submit the work in pairs. Only one submission for each pair.
- Your submission should include a single zip file containing:
  - A single ipython notebook (.ipynb) file that includes all the code and outputs.
  - A brief report (1-2 pages) in a pdf format that explains your approach and findings.
- The submitted file should follow the naming convention:
  - 3945\_HW##\_XXX\_YYY.zipWhere:
  - ## is the assignment number
  - XXX and YYY are your student numbers (IDs)For example: 3945\_HW1\_123456789\_987654321.zip
- Make sure to run your notebook from start to finish before submitting to ensure that it runs without errors.
- You may use external libraries. Specify all required libraries in a proper manner.
- Grading will be based on correctness, elegance of solution, and style (comments, naming conventions, etc.)
- Your report should be clear, coherent, and concise.
- All figure and plots should include captions, labels and data units. Pay attention to data visualization guidelines.
- Make sure to use correct ML methodologies and justify your selections (split the data to train/test, tune hyperparameters, report relevant performance measures).