Neural Networks and Deep Learning (2025)

Project 1:

Clustering with SOM, Classification with SLFN

Due date: 22th Faravardin 1404 [23:59]

Introduction

In this homework, you will explore the differences between unsupervised learning (clustering) and supervised learning (classification) using neural networks. You will first apply a Self-Organizing Map (SOM) to cluster the digit dataset, performing both clustering and classification and analyze its performance. Then, you will implement a Single Layer Feedforward Network (SLFN) from scratch to classify the titanic dataset.

Task 1: Clustering Digits Dataset with SOM (Grid Size 4x4 and 20x20)

Dataset: Digits dataset

The dataset contains 1,797 grayscale images of size 8x8 pixels, with 10 classes representing digits 0-9. You will use the digits dataset from sklearn.datasets, which can be imported using this code:

```
from sklearn.datasets import load_digits
digits = load_digits()
```

Steps to follow:

1. Prepare the dataset.
2. Implement and train a Self-Organizing Map (SOM) using the digits dataset.
3. You can use the **MiniSom** library or another approved SOM package.
4. Train the SOM with the following configuration:
   - **Grid size: 4x4 (16 neurons)**
   - **Grid size: 20x20 (400 neurons)**
5. Visualize the clustering results:
   - **Display** a few sample images mapped to each neuron (cluster).
   - Identify neurons with no images (dead neurons) and **explain** why.
   - **Interpret** whether the images in each neuron show meaningful clusters.
   - **Hit Maps** (density maps showing the number of data points per neuron)
6. Calculate the **quantization error** to evaluate how well the SOM represents the data.
7. **Compare** the clustering results
   - Are clusters more refined?
   - Are there more dead neurons?

Task 2: Classification of Digits Dataset Using SOM

Steps to follow:

1. According to previous task, assign a **label** to each neuron using **majority voting** from the training data mapped to it.
2. Classify the test set:

   - For each test image, find its **BMU** (Best Matching Unit).
   - Assign the label of the BMU as the prediction.

3. Evaluate the classification performance:

    - Calculate **accuracy** on the test set.
    - Provide a **confusion matrix**.
    - Calculate **precision, recall, and F1-score** for each class.

4. Visualize the neuron class map:

    - Create a **color-coded map** where each neuron is colored according to its assigned class label.
    - Mark dead neurons with a neutral color.

5. **Compare** the **Classification** results

Task 3: Classification with Single Layer Feedforward Network (SLFN)

Rules:

- You **must** implement SLFN **from scratch** using only **NumPy** and **Pandas** if you are using **python**.
- **NO TensorFlow/PyTorch** or **other deep learning libraries allowed**.

Dataset: Titanic dataset

The Titanic dataset consists of 892 samples of passengers. The dataset includes details about each passenger, such as age, gender, ticket class, and whether they survived. Your task is to **predict whether a passenger survived or not** based on their features using a **SLFN**.
You can download the dataset here: DATASET

Steps to follow:

1. Prepare the dataset and **discussion** on how you preprocess the dataset.
2. Implement SLFN from scratch:
    - Initialize random weights and biases.
    - Implement the forward pass using **a single hidden** layer (**ReLU** activation).
    - Implement the **sigmoid** output layer for classification.
    - Compute **cross-entropy loss**.
    - Implement **backpropagation** to update weights (using **gradient descent**).
    - Train the model for a fixed number of epochs.
3. Evaluate the model:
4. Compute **evaluation metrics (Accuracy, Recall, Precision, and F1-score)** on the test set.
5. Compute **confusion matrix**.
6. Plot the **loss curve over training epochs**.
7. Plot the **AUC-ROC curve**.

Notes:

- Allowed programming languages: **Python**, **MATLAB**
- You can use pre-defined libraries for this assignment, except for SLFN implementation.
- Any sign of cheating would result in a **zero** grade for this assignment.
- You should upload your submissions at: Link

All of the files should be in a ZIP file named in this format:
"FirstNameFamilyName-SudentNumber.zip"
Ex: "AmirZamani-4033040.zip"

- Your reports should be in a PDF file including: key points of your implementation, explanation of your chosen approach, reports of your final results and answers of assignment questions (if given)