Artificial Intelligence and Expert Systems
Assignment 2

Instructors:
Dr. Esmaeil Najafi
Armin Ghanbarzadeh
Reza Behbahani Nezhad

Due Date: 31 Farvardin, 1403

Introduction

Human activity recognition (HAR) refers to the task of automatically identifying and classifying human activities based on sensor data collected from wearable devices or other sensors. The goal of HAR is to build machine learning models that can accurately recognize different activities performed by humans, such as walking, running, sitting, standing, or even more complex activities such as climbing stairs or dancing. HAR has important applications in various fields, including healthcare, sports, and smart homes. By using HAR technology, healthcare providers can monitor the physical activity of patients and assess their recovery progress. Sports professionals can use HAR to track athletes' performance and prevent injuries. In smart homes, HAR can be used to control home appliances and adapt the environment to the user's activities and preferences.



The dataset was obtained using the integrated sensors of a smartphone, specifically the Accelerometer, Magnetometer, and Gyroscope. From these sensors, four types of parameter readings were recorded in the X, Y, and Z directions. The X_{acc} parameter represents acceleration in the X direction, X_{AV} denotes angular velocity in the X direction, X_{mf} indicates the magnetic field in the X direction, and X_{orien} represents the orientation reading in the X direction. Similar parameters were recorded for the Y and Z directions.

The dataset consists of nine distinct classes, with each class represented by a single file. To address class imbalance issues, each file contains 403,605 data points. Data collection was conducted at a sampling frequency of 100 Hz, with a duration of approximately 1.12 hours per class. These details of the dataset are crucial for researchers and practitioners who seek to work with the data and require a clear understanding of its characteristic.

More information on the dataset can be found in this Kaggle link.

Experiment 1: Multi Layer Perceptron (MLP)

With the dataset, perform the following steps:

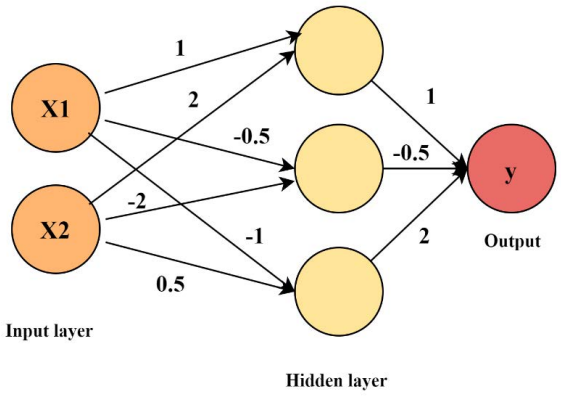
(a) What would be the output of a Multi Layer Perceptron (MLP) if only linear activation functions are used, and what problems can it cause? Explain the impact of activation functions on the performance of an MLP network.

(b) Plot and explain Leaky-Relu, PReLU, RReLU, ELU, and SELU activation functions.

(c) Briefly explain weight initialization and compare three methods of weight initialization: (I) assigning equal weights zero to all neurons, (II) sampling Gaussian function, and (III) Xavier initialization.

(d) For a simple regression task, the simple MLP network shown below is used with linear activation functions for hidden layers and the output layer. Using MSE loss function, calculate the partial derivatives of the loss function with respect to the biases and weights of the network. Train the network for one epoch and update the network weights and biases. Consider the biases equal to zero at first, and X1 =3, X2=2, y=5

Note: You should perform all calculations by hand (without using any code).



(e) Explain what dropout is? What are the differences in its implementation during the training and testing stages of a deep neural network.

(f) Define and explain the concept of batch normalization in the context of machine learning and neural networks.

(g) Consider the Dataset provided and Implement a network with at least 4 hidden layers, and then train the network again using dense layers followed by dropout and batch normalization.

(h) Compare the results and write your own conclusion. Does the network become overfit, underfit, or train normally? Give reasons for your answer.

(i) What are the popular ways to prevent an MLP from becoming overfit? Mention two ways.

(j) What are the hyperparameters of your networks. What are some ways to tune these hyperparameters efficiently? Mention some popular ways such as mesh grid search and random search.

Experiment 2: Recurrent Neural Network (RNN)

(a) Explain the basic architecture of an RNN and how it is different from a feedforward neural network. What are the advantages and disadvantages of using an RNN?

(b) What is the problem of vanishing and exploding gradients in RNNs? How can this problem be mitigated?

(c) Explain the difference between simple RNNs, LSTM, and GRU. In what situations might each type of RNN be most appropriate?

(d) Explain how RNNs are trained using backpropagation through time (BPTT). What are some of the challenges associated with training RNNs, and how can these challenges be addressed?

(e) What happens to the gradient if you backpropagate through a long sequence?

(f) Train a network using simple RNN cells

(g) Train a network using LSTM Cells or GRU (you can use both)

(h) Train a network using bidirectional LSTM or GRU (you can use both)

(i) Tabulate the results, compare them and write your findings.

(j) What are the hyperparameters of your networks. What are some ways to tune these hyperparameters efficiently? Mention some popular ways such as mesh grid search and random search.

Bonus Question: Exploring Attention Mechanisms

As you delve deeper into neural networks, you'll encounter various architectures and techniques that push the boundaries of machine learning. One such advancement is the concept of attention mechanisms or transformers.Your task for this bonus question is to explore the paper [Attention is all you need] and provide a simplified explanation of what they are and how they work. Your explanation should cover the following points:

(a) Definition of Attention Mechanisms: What are attention mechanisms in the context of neural networks? How do they differ from traditional feed-forward or recurrent architectures?

(b) Functionality: How do attention mechanisms enable neural networks to prioritize certain parts of input data? What problem do they solve in machine learning tasks?

(c) Implementation: Briefly describe how attention mechanisms are typically implemented in neural network architectures. Are there different types of attention mechanisms?

(d) Application: Discuss the significance of attention mechanisms in advancing the capabilities of neural networks. Provide examples of real-world applications where attention mechanisms have made a difference.

Remember, this is a bonus question, so feel free to dive as deep as you like into the research paper and subject!

Important Notes

- Please submit your homework assignment as a zip file containing **(a)** a PDF report (analysis, results, methodology, ...), and **(b)** code files necessary to reproduce your results.
- All grading will be based on the content of the PDF report. Make sure to include and explain your code in the report.
- Please make sure to submit your solutions by the due date. No late submissions will be accepted.
- Assignments are to be completed individually. Any similarities between assignments will be subject to reduced grades.
- If you have any questions, feel free to ask.

Good Luck!