VU Machine Learning

Summer semester 2025

Exercise 2

Nysret Musliu (nysret.musliu@tuwien.ac.at)

Patrick Styll (patrick.styll@tuwien.ac.at)

- Groups of 3 students

- Implement techniques for classification

- Compare to existing techniques

- Submit the source code

- Prepare around 10 slides

- Individual discussions (20 min) for each group (all members must be present)

- Submission: May 26th

- Discussions: after May 26th (slots will be available in tuwel)

- Implement a framework for neural networks (NN)

- Implement the entire network from scratch, including forward and backward propagation methods

- Implement the same network using PyTorch (use standard functions)

- Use an LLM tool to implement the same NN from scratch. Discuss the differences between your custom implementation and the one provided by the LLM tool

- Investigate and experiment with different NN configurations, including 2–3 activation functions, different numbers of layers, and different numbers of nodes per layer. You can implement a simple method based on grid search to automatically find good values for these hyperparameters

- Write a function that calculates the total number of learnable parameters and the virtual RAM used by an instantiated NN

- Your algorithm should work for classification datasets that include nominal and/or numeric attributes (features)

- You should implement the methods from scratch

- Please do not use any part of existing code

- You can use existing code/functions for general parts like

    – Mathematical calculations (derivatives…), code for reading the input and testing the algorithm (cross- validation, performance metrics…)

- Pick 2 classification data sets

  – Data sets from the previous assignments can be used

  – Or other data sets from UCI ML Repository, Kaggle…

- Should have different characteristics

  – number of samples – small vs. large

  – number of dimensions – low vs. high dimensional

- Pre-process the data set if needed (scaling, missing values …).

  – A extensive pre-processing is not the focus of this exercise

Comparison

- Compare your implemented techniques on selected datasets with

  - PyTorch implementation

  - Existing NN implementation (sklearn implementation, …)

    - You can use the default parameters for the existing technique

- Conclusions

  - How efficient is your algorithm and what is the total number of learnable parameters and virtual RAM used by your final NN

  - Performance of your algorithm regarding performance metrics for classification

  - Findings from code comparison with the LLM tool and Pytorch implementation

  - Other findings and lessons learned

A zip file with

- **Source code:**

  – For your implementation from scratch you can use any programming language: Python (recommended), Matlab, R…

  – Provide the information for the packages needed to run you code

- **Slides**

  – Around 10 slides

  – No report needed

- Submission deadline: May 26th, 23:00

- A brief discussion of the implementations

- Discussion of differences and similarities between your implementation and the Pytorch/LLM implementations

- Comparison with the existing implementation

- Discussion of experimental results

- Conclusions and lessons learned

Individual discussions

- Length of discussion: 20 minutes

- You will have questions about

  - Source code

  - Techniques

  - Comparison with the existing techniques

  - Lesson learned

- All members of the group should be able to explain the code/experiments

  - The evaluation will be based on your code, discussion, comparison, and conclusions/lessons learned

- **Note that the main aim of this assignment is to ensure that every student is able to implement a neural network from scratch, including the details of forward and backward propagation**