Exam session 02 September 2025

509486 - Machine Learning, Artificial Neural Networks and Deep Learning - Part 2

[L-31] Artificial Intelligence

—

We are given a dataset consisting of $n = 5000$ authors, categorized into three initial classes:

- Authors whose primary research field is *Database* (label value **1**);
- Authors whose primary research field is not *Database* (label value **-1**);
- Authors whose primary research field is unknown (label value **0**);

There are 750 authors with unknown primary research field, and among the remaining ones, 1900 do research in the Database domain, and 2350 in domains other than Database.

Additionally, information about authors, such as the distribution of terms, the topics, and the co-authorship patterns relevant to their research is cast into an author similarity matrix $\boldsymbol{W} \in [0, 1]^{n \times n}$, where

- $W_{ij} \in [0, 1]$ indicates a pairwise similarity between authors i and j,
- $W_{ij} = 1$ denotes the maximum possible similarity,
- $\boldsymbol{W}$ is symmetric, obviously, that is $W_{ij} = W_{ji}$ for any $i \neq j$.

Design a neural architecture capable of predicting the binary class labels of the authors, whose primary research field is unknown, by leveraging both the available labels and the similarity relationships between authors. Assume that the true labels associated to 0-labelled authors will be available during the implementation of the proposed solution.

Take 5 minutes to think about the problem, then provide a clear answer to each of the following points. Please, write in a READABLE way.

1. MODEL CHOICE. Which architecture do you consider the most appropriate for this task, among those covered in our course, and WHY;
2. INPUT. How to (if) preprocess input data;
3. MODEL CHARACTERISTICS.
    a Model composition: overall (graphical) outline;
    b Model formal description (mathematical properties and rationale of the solution);
    c How the labels for the authors with unknown primary research field are inferred;
    e Are there parameters/hyperparameters to be learnt/tuned?
4. MODEL EVALUATION. Describe how would you evaluate the quality of model predictions.

HOW TO ANSWER

- Motivate your choices. Provide technical/mathematical descriptions when suitable. Writing more does not necessarily imply a higher evaluation, in general. Be precise.
- Dedicate an answer to each of the points above (1-4), **maintaining the same numbering format, even for subitems**.
- Please leave SPACE after each point. Insert line breaks to help readability.

During the exam, the use of electronic devices, including mobile phones, tablets, smartwatches, etc., is strictly prohibited. Any detected instance of misconduct will be appropriately sanctioned and recorded in the student's academic record.

You will have time till September 9th, 23:59, to upload the implementation of your solution in a Colab notebook. The notebook should totally adhere to the proposed solution. On the web page of the course you will find a link with the instructions on how to submit your solution. At the same link, after the exam, you will also find the abovementioned data to be used for the implementation.

If no file is uploaded, the exam is considered as rejected. In case of problems in uploading your solution, write for the teacher BEFORE the deadline for submitting.