Exam session 19 June 2025

509486 - Machine Learning, Artificial Neural Networks and Deep Learning

[L-31] Artificial Intelligence

—

We are given a dataset containing reviews about some hotels, written down by 13772 visitors. The following fields are available:

- *Hotel_Name*, string;

- *Hotel_Address*, text;

- *Review_Date*, the date of the review, in the format *month*/day/*year*;

- *Hotel_number_reviews*, integer, the total number of reviews available for the hotel;

- *Reviewer_Nationality*, string;

- *Reviewer_numer_reviews*, integer, the total number of reviews the reviewer has already made;

- *Review*, text, the whole text of the visitor's review. The maximum number of words in a review is 400;

- *Review_Score*, float, the score given to the hotel. It is a value in [0,10], with 2.5 and 10 as minimum and maximum values in the available data;

- *Review_Type*, string, if the review is good or bad. Two possible values, *Bad_review* and *Good_review*. Their total number is, in the order, 6888 and 6884.

Design a deep neural network able to predict the review type and/or the review score.

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

1. MODEL. Which architecture do you consider the most appropriate for this task, and WHY;

2. INPUT

    a How to (if) preprocess input data;

    b Which is, after the preprocessing step, the input of the model: type, shape, value domain;

3. OUTPUT. How would you design the output layer and why;

4. LOSS. Which loss function and labels would you use to train your model and why;

5. MODEL CONFIGURATION

    a Model composition (composition of layers, regardless their number, or their dimension)

    b How do you intend to configure your model to optimize your choice, mention a few hyperparameters most relevant in your opinion.

6. MODEL EVALUATION. How would you assess (in which setting) the generalization capabilities of the model on unseen data?

HOW TO ANSWER: briefly motivate your choices, but provide enough details. Writing more does not necessarily imply a higher evaluation, in general. Be precise.

Dedicate an answer to each of the points above (1-6), **maintaining the same numbering format, even for subitems**.

Please leave SPACE between answers.

You will have time till June 28th, 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 above-mentioned deadline for delivering.