ΥΣ19 Artificial Intelligence II (Deep Learning for Natural Language Processing)
Fall Semester 2022-2023
Homework 2
25% of the course mark
Announced: November 30, 2022 Due: December 20, 2022 before 23:59)

In this homework you have to develop a sentiment classifier for the dataset `imdb-reviews.csv` of Homework 1 using feed forward neural networks.

In this homework, you should use the machine learning framework PyTorch (`https://pytorch.org/`).

Before you do the homework, make sure that you have studied the relevant slides of the course (PDF files "Perceptrons", "Backpropagation", "Training DNNs", "Word Vectors" and "word2vec") or any other relevant literature you may find useful.

The inputs to your model must be GloVe word embeddings (`https://nlp.stanford.edu/projects/glove/`).

As in Homework 1, it is your responsibility to choose all the details of developing a good neural network model for the problem (e.g., see page 104 of slides "Backpropagation" regarding which hyperparameters you need to tune). It is expected that you will experiment with more than one neural network architecture to choose the best for the problem.

You should plot learning curves that show that your models are not overfitting or underfitting.

You should evaluate your classifier using *precision, recall* and *F-measure* like you did in Homework 1. In this homework, you should also plot ROC curves (`https://en.wikipedia.org/wiki/Receiver_operating_characteristic`).

Your code should be written in a way that your models can be evaluated on the test set by simply passing the path of the test file to a specific variable.

You should hand in:

1. A PDF document with a detailed explanation of your solution including citations to relevant literature that you might have used in developing your solutions. If you use LaTeX for your document, you will get a bonus 5%. In this case, you have to hand-in the LaTeX source files too.

2. One or more Colab notebooks (`ipynb` files using `https://colab.research.google.com/`) containing your code.