Hands-On Assignment

Deep Learning
880008-M-6

Juan Sebastian Olier

Spring, 2020
Submission deadline: March 15th, 2020

Description

This is a group assignment. Groups should be of three people. For this assignment, you will participate in a shared task, which consists of a classification problem. You will need to submit your predictions on a given test set, as well as a report describing your solutions and their connection to the scientific literature, and the code you used to generate such predictions. This assignment is worth 30% of your course grade. The assignment grade will be based on the quality of your work as judged by the instructor based on your report and code. Additionally, you will get a bonus based on your ranking on a leader board of the shared task. Specifically:

- if your rank first, you will receive bonus 2 points;
- if your score is no better than a provided baseline you will receive no bonus;
- for intermediate ranks the bonus points will be linearly interpolated. The performance of the baseline solution will be shown on Codalab.

All the information about the Codalab submission will be specified on a separated document later on.

Passing the assignment is not mandatory to pass the course, but it is highly advisable. As it is not compulsory, there will be no resit. The exam may include questions that might be easier to answer if you have worked on it.

1 Task - Classification of EEG

Analyzing sequential data or time series is a very relevant and explored task in Deep Learning. This kind of data appears in many domains and different formats; for example, stock prices, videos and electrophysiological signals.

In this task, you will be working with Electroencephalography (EEG) or brain waves. The data set consists of EEG signals of different subjects while sleeping. There are different stages of sleep characterized by specific kinds of EEG signals. Each sleep stage is then a period during which the EEG signals present specific features or patterns, such as particular frequencies as can be seen in 1 (The frequency bands in the table are averages, and other sources might define them slightly differently, but these are representative) - For a more comprehensive description, please go to https://www.ncbi.nlm.nih.gov/books/NBK10996/ You do not need to use all of the information regarding the definition of stages unless you find it useful or want to go more in-depth about the problem.

The task you have to address is to classify EEG signals into sleep stages.

1.1 Data set

The data-set consists of EEG sequences of 3000-time steps each and coming from two electrode locations on the head (Fpz-Cz and Pz-Oz) sampled at 100 Hz. That means that each sample contains two signals of 3000 samples and that those samples correspond to 30 seconds of recording.

The labels that come along with the data specify six stages labelling them with corresponding numbers as specified in table 1:

Label	Stage	Typical Frequencies (Hz)
0	R	15-30
1	1	4-8
2	2	8-15
3	3	1-4
4	4	0.5-2
5	W	15 - 50

Table 1: Caption

W corresponds to the Wake stage, and R to REM sleep also called rapid eye movement, and most commonly known as the dreaming stage.

Each sequence in the data set contains only one stage, which is specified by the corresponding label.

Your final model should be evaluated on a given unlabeled Test set that will be provided later.

The data set is presented in two different formats, Raw signals and Spectrograms; To solve the problem, you can use either or both formats, there is no restriction.

1.1.1 Raw signals

The file *'Data_Raw_signals.pkl'* contains the sequences and the corresponding labels as two array [sequences, labels].

1.1.2 Spectrograms

The file *'Data_Spectrograms.pkl'* contains the spectrograms of the sequences and the corresponding labels as two array [spectrograms, labels].

A spectrogram is a visual representation of the spectrum of frequencies of a signal as it varies with time. So a spectrogram is a 2D array, where one axis represents frequencies, and the other represents time. https://en.wikipedia.org/wiki/Spectrogram

In the file *'Data_Spectrograms.pkl'* the spectrograms have a size 100 by 30 for each signal, and they represent the same 3000-time steps EEG sequences as in the raw data. That is, for each sequence in the raw data file, there is a corresponding spectrogram.

The spectrograms in *'Data_Spectrograms.pkl'* represent the frequencies of the signals in steps of 0.5Hz between 0.5 and 50 Hz (Hence 100). Such frequencies correspond to the spectral information in time windows of size 100-time steps each; thus, for a sequence of 3000-time steps, there are 30 windows (hence spectrogram size: 100 by 30, frequencies by the number of windows).

2 Important dates and deliverables

2.1 Report

A one-page report should be submitted by March 15th, 2020. The report should include the following:

- Title including names and student numbers.
- A clear connection to scientific literature.
- Deep Learning architectures used. This description must be centred around a diagram depicting your design.
- Brief description of your experiments, including training, hyperparameters, and optimization.
- Results with a table, a graph or both.
- Discussion of the performance of your solution and how it relates to the literature used.
- The name of the account under which you submit your results to Codalab.

Your report should be a PDF document, with a single page of content, and optionally additional pages for references and appendices. Note that the content page needs to be self-contained, and the appendices should only contain auxiliary material.

2.1.1 Code

Your code should be a plain Python script which can be run to generate your predictions. You do not need to include the training data or trained model.

2.1.2 Submission format

Put the report and the code in a single zip file named with your group ID, e.g. group_1.zip, and submit it to the Canvas assignment.

2.2 Performance

In addition you will need to submit a file with the predictions to the competition server. The competition will be hosted on https://competitions.codalab.org. You will need a Codalab account for the group: indicate the name of this account in your report. There will be a separate document on the submission to Codalab with additional details.