

Assignment 2 – Practical Deep Learning Workshop

Submission by
may 11th

* this assignment can be done in teams of up to 2 students

Time series – register to the <https://www.kaggle.com/competitions/tlvmc-parkinsons-freezing-gait-prediction> competition .

Your task is to **best classify 3 types of activities** conducted by the Parkinson's Disease patients (PD)

- 1) Present an exploratory data analysis of the dataset. 15 pts
 - a. You can refer to the previous assignment, but also think what else can be interesting about this data.
 - i. Where is the data taken from? Is it homogeneous or does it vary in some way?
 - ii. How was the data labeled?
 - iii. Should all the labels be treated equally? Are some more important? Are all validated?
 - iv. How many subjects are in the data? How was the competition data split to train/test?
 - b. Show some plots that explain the task at hand, is it classification? or regression? Are you trying to predict a future event, or predict some additional information of a past event?
 - c. state two **self-supervised** tasks as we have seen in class to pretrain your models for next sections on
- 2) Form a neural network graph based on the components we used in the walkthroughs in class. 75 pts
 - a. Decide your validation strategy for training your model
take into account your insights as to how the data was curated and partitioned
 - b. Create a naïve baseline solution and calculate train and validation score for that solution. This could be using last known value, or the class distribution for each category as naïve prediction.
(for those of you who wish to extend this section – consider tsflex/rockets/sktime/tslearn libraries)
 - c. Fit a classical machine learning model to some features you extract from the data to get a **better** and **solid benchmark** for the neural network model. Think which features will be useful for such a model
 - d. Construct 2 different neural network models and fit them to the data. analyze the results (Use visualizations to present your loss and other metrics you find relevant, show examples for good and bad classification with high probability, and also present examples of uncertain predictions.
Compare the results you got on the training data vs. your results for the validation/test data)
 - i. 1D-CNN
 - ii. LSTM/GRU single/bi-directional
 - e. Pretrain your model on one of the tasks you suggested in 1c or adjust an external pretrained TS model and fine tune the trained model to the data. compare the results you got to previous sections (c-d)
 - f. Try to figure out where & why is the model doing well and suggest at least 3 ways to improve the results
 - g. Prioritize the list of suggestions for improvements and implement the first 2 suggestions, repeat section d. and present a table of all your results (you can use a tool like Neptune.ml to record your results)
- 3) Write a report that summarizes your research. 10 pts

Think what you would like to emphasize and what is of lesser importance present your analysis and report

5 pts bonus to overall course grade will be granted to each group member of a group that passes 0.4 mAP until submission date

* You are free to use any external data source for the pre-training task