Master Internship/Graduation Project:
Predicting classification accuracy without labels

Contact: Iris Huijben (i.a.m.huijben@tue.nl)

Background:
Sleep is a vital part of a human being's day in order to keep the body in a well-functioning state and 'clean up' unnecessary memories in the brain. To measure someone's sleep, a polysomnography (PSG) recording is used. Such a measurement includes (among others) electroencephalography (EEG), chin electromyography (EMG) and electrooculography (EOG) data during sleep. The American Academy of Sleep Medicine (AASM) [1] distinguishes different states through which a sleeping brain transitions during the night: rapid eye movement (REM) sleep, non-REM sleep (subdivided into N1, N2, and N3), and wakefulness. Given a PSG recording, a sleep expert labels each window of 30 seconds with one of the five possible states to create a hypnogram; a visual representation of assigned sleep stages over the full night. This manual labelling process is time-consuming, and could be improved by automating this process.

Problem description:
Many machine learning algorithms have already been developed for training a supervised classification model that performs automatic sleep staging. In real-life situations, it is, however, thinkable that large amounts of recordings are available, while being unlabelled. This hampers training a supervised classifier, which resorts us to unsupervised training. Contrastive Predictive Coding [2] has earlier been proposed as an unsupervised learning technique to pretrain a feature extractor on large amounts of unlabelled sleep data, after which a supervised classifier can be trained a small labelled dataset [3].
When running new unseen (test) data to the trained model, to predict the sleep stages, it is hard to know whether predictions can be trusted, since a ground-truth label is not available.
The question that now arises is, can we use the loss function of the unsupervised pre-training (or metrics computed from this loss function) as an indication for how well the classification performance will be on this window?

Methods:
The goal of this project is to investigate whether we can predict, without access to label information, how easy it is to correctly classify a data point. If we can do this successfully, it can provide the user with information regarding trustworthiness of the model prediction.

Requirements:
- Knowledge in deep learning, for example from course 5LSL0 (or similar courses).
- Python
- Experience in Pytorch (or Tensorflow)

Duration: This project can be done as 3-month internship or a full graduation project. Though, for an internship, having already high proficiency in python and Pytorch programming is essential in order to speed up the start-up phase.

References:
[1] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. Marcus, B. V. Vaughn et al., "The AASM manual for the scoring of sleep and associated events," Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine, vol. 176, p. 2012.

[2] Oord, A. V. D., Li, Y., & Vinyals, O. (2018). Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

[3] Banville, H., Wood, S. U., Aimone, C., Engemann, D. A., & Gramfort, A. (2022). Robust learning from corrupted EEG with dynamic spatial filtering. *NeuroImage*, *251*, 118994.