1 Introduction

Presentation The goal of this project is the prediction of a subject's gender from its brain rhythms. The dataset provided by Dreem for this project is divided into a *training set* and a *test set*. The training set and the test set both contains EEG measurement for 946 subjects. On each of these subjects, the EEG signal has been recorded on 7 channels on 40 time intervals of 2 seconds. The sampling frequency of the signal is 250 Hz. As a consequence, each subject is associated with an input array of size $(40, 7, 500)$. The recordings were made during the night.

Notation In the following, male subjects will be associated with the class 0 and female subjects will be assigned class 1.

Previous work The challenge webpage was presented with a reference article [?]. In this work, the author applied a deep-learning model to the prediction of the gender from EEG recordings. Using a convolutional neural network, the authors achieved 81% of accuracy.

Data balancing The repartition of genders in the training set is 737 men and 209 women, i.e male subjects represent 78% of the training set. As a consequence, a trivial classifier returning 0 will have 78% accuracy on the training step. It is natural to ask whether or not the test set is also imbalanced. Since, we do not have access to the true classes of the test set, we simply submitted a solution on the challenge webpage where all the predictions where male, and obtained and accuracy of around 77.5%. Thus the test set has this 77.5% of male subjects and is quite imbalanced like the training set.
There are two possibilites to counter the imbalance : we can either oversample the class 1 to adjust the class distribution or assign different weights to each class.

2 Deep learning-based classification

In this section, we develop the deep learning pipeline used for the gender classification.

The original article on which the challenge is based proposed [4] describes an architecture based on a previous architecture designed by Krizhevsky et al. to work on ImageNet [2].

2.1 Braindecode library

We used the Python library [?] specialized in classification of EEG recordings using deep-learning, and is built on top of the Pytorch library.

Shallow neural network

Deep neural network

Visualizing the network parameters

3   Classification with linear models

References

[1] Schirrmeister et al. Deep learning with convolutional neural networks for brain mapping and decoding of movement-related information from the human EEG. *CoRR*, 2017.

[2] Sutskever Hinton Krizhevsky. Imagenet classification with deep convolutional neural networks. *Neural Information Processing Systems*, 25, 2012.

[3] Tasali Hofmann Van Cauter Latta, Leproult. Sex differences in delta and alpha eeg activities in older adults. *Sleep*, 2006.

[4] Arns van Putten, Olbrich. Predicting sex from brain rhythms with deep learning. *Scientific Reports*, 2018.