# Short description of chosen topic for assignment 1

Kristof Furuglyas
Data Mining and Visualization

October 25, 2019

## 1 Motivation

My topic is about somehow applying machine learning and data mining methods on EEG data. Since I have a lot of data and different measures, my main goal is to find such a method that can either group these measures or at least validate the theories we already have.

## 2 Basics

Mismatch negativity is a well-known phenomena that is present in the brain when a deviant signal is present after many standard has been shown. By somewhat trying to learn the incoming signal and/or predicting it, the local connections (rather neural activities) form in such a way that the measures EEG signal drops down i.e. the brain gets used to the same input.

Let us say, the brain gets used to stimulus A, and after a lot of A (>20 but may vary) we present stimulus B. The response we get is not the same as if we just randomly shuffled the responses (equiprobable experiment). Two main ideas lay beneath the surface; first that the difference comes from the fact that the received signal is not the same as the predicted one, second is that when processing the signal other areas of the brain may join into the decomposition i.e. we do not measure the same population of neurons.

However, in our case we have two types of stimulus: visual and auditory. This setup allows us to measure whether conditional mismatch negativity exists. Combining the to types of stimulus with the two kinds (horizontal or vertical grating to the eyes and low or high pitch noise to the ears) gives us all in all 4 types of stimulus, let us call them A-D. From this one can identify the mismatches: only visual, only auditory or bimodal.

## 3 Data availability

The experiment has been done many type on both anaesthetized and awake mice in the visual and auditory cortices of the brain by Holland biologists. During my job I always

worked on only one anaesthetized mouse, but I can access the data easily if needed. For the main experiment the stimulus lasted for 500 ms with 1.5 sec of ITI (inter-time interval). One session (when one was the standard) meant approximately 600 stimulus with a $32kHz$ of sampling frequency, all in all I have the access to terabytes of raw EEG data.

# 4 Goals

As it was mentioned in the introduction data clustering can be useful and maybe dimension reduction. For the latter TSNE and UMAP are available, yet, I have no precise idea about the former but maybe Andrew NG's publications (which was mentioned by Bálint) might help – use of ML in neurology. All the results I have so far can be tested both by supervised and unsupervised learning, both in time and frequency domain.

# 5 Tools

I will definitely use Python in Jupyter notebook. The packages probably will be the same as on the lectures. No other external tool is in plan.

# 6 Errors and sources

Due to the limited resource of counting capacity, many problems may arise with memory-maintaining. Furthermore, the creation of any type of table is problematic, since there is no concrete perspective on how to represent a signal. Numerous tables probably will be dummy tables.

# 7 Expectations

Any kind of help with machine learning would be highly appreciated. Deeper meanings (in a biological sense) of the results may go beyond this project's borders, but it is still a great practice in machine learning, indeed.