Homework 2. Multivariate Time Series and Reservoir Computing

The goal of this homework is to have a first hands on contact with multivariate autoregressive models and with echo state (reservoir computing) networks as tools that aim to capture the temporal, interdepent structure of multidimensional temporal series.

To this end, we will use a set of EEG data recorded during a psychophysical experiment aimed at testing a hypothesis of the influence of motivation onto movement.

Experimental Description

We will use a dataset from an experiment for the study on decision-making between movements of opposite motor cost (FIG 1A). We asked our participants to perform this task under several conditions of different social pressure (FIG 1B). This manipulation was aim at inducing different motivated states, which we will use to measure by means of electro-encephalographic recordings.

A single participant takes part in each session. Each session is composed of several trials, to gather the neural and behavioural data we will analyse next. Each trial follows the same timeline: it first starts with the presentation of a geometrical distribution (blue rectangles) from an origin cue (a small red circle) presented at the centre of the computer screen (FIG 1A). Once the stimuli have been presented, the volunteer must choose a target and make a planar reaching movement, while trying to touch the rectangle of his/her choice which was synchronized as a function of the finger position. At each trial, we recorded the index finger trajectories as a response of the presentation of the stimuli by means of a 3D tracking device, as well as electro-encephalographic signals, which provide us with a metric of the brain state at the beginning of the trial.

It is well known that the cost of each movement exerts an influence on each trajectory and choice between trajectories. Our hypothesis is that our motivated state also influences our decisions between movements and/or the specifics of each movement we perform. To assess that, we measured how often the volunteer chose the right/left rectangle and analysed the speed, movement and error rate.

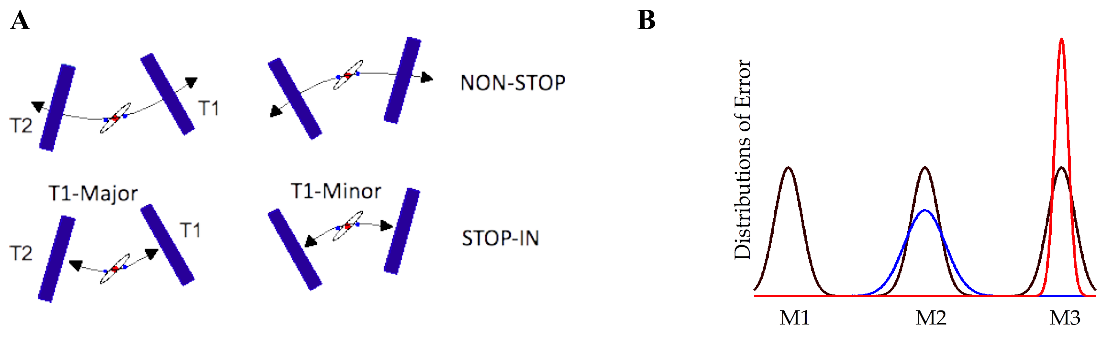

Figure 1. A. Presentation of geometrical schemas for decision-making. At each schema, there is two goals (T1 on the right or T2 on the left), a red origin cue. All trajectories depart from that origin cue. To assess cost difference we designed two T1-Major schemas (the movement towards the right implies a lesser cost, than the one towards the left), and T1-Minor (reversed costs with respect to the T1-Major arrangement). **B.** Manipulation of the internal motivation of the participant. Distributions of error of the participant (black) and of their partners (blue --- lesser skill, and red – higher skill).

Although the goal choice (rectangle) is free, the goal of each movement is that of winning points as by being more precise by arriving at the centre of the rectangle long side. The more central is our arrival, the more points we will win, the further off is our trajectory, the lower is the outcome. Furthermore, upon movement completion, we showed the subject a green error bar [0-100%], indicating the amount of points earned in that trial.

We used social pressure to manipulate the participant's motivated state. In this context, the manipulation of the motivational state consists of simulating the presence of a partner player with a skill different from that of the participant. In summary, we create avatars to induce three potential motivated states:

- *State 0 (Motivation 0),* play alone (most relax).
- *State 1 (Motivation 1)*, Play with a player of a lesser skill.
- *State 2 (Motivation 2)*, Play with a player of a higher skill level.

Furthermore, every time our participant finishes a trial by movement execution, we show him/her the amount of points he/she won, contingent on the precision upon target arrival. We also show the points of the partner on that specific trial (when playing with a partner). To prevent that the subject starts a competition against the partner, we instructed him/her that the purpose of the partner's presence is not competition, but rather keeping company to the participant during task performance.

Data Recordings
In addition to trajectories, which report the decision and assess movement precision, we also record electro-encephalograms, which allow us to extract a metric of brain activity during decision-making for each motivated state.

DATA ANALYSIS
The part related to movement analyses is provided (no need to care about that). FIG 2 (left) shows three typical trajectories in each of the three motivated states, in the centre it shows the precision error of our participant in each of the three motivated states. In the centre, it shows the precision error of our subject in each or the thee motivated states (1,2,3), and the graph in the centre shows the frequency with which it selected the right target as a function of the motivated state.

The dataset is available at the following link:

https://ubarcelona-my.sharepoint.com/:f:/g/personal/ignasi_cos_ub_edu/EpKJUjB68ChCnJGOPYvgx9QBcnweW8xpZ1TdZDtITmVKQg?e=VRKldp

YOUR TASK

Your dataset consists of a 42 dimensional time series of electroencephalographic activity extending for over 1200ms, and organized in blocks of trials 432 trials of one of three types (Class 0, class 1, class2). In summary, your dataset has 1200ms x 42 channels x 432 trial repetitions x 3 Classes.

Your first task consists of classifying the data in a two-step process:

1. Using a multivariate autoregressive model. Choose the proper autoregressive model from all the ones you recorded from, fit that model to the time series of each class. You should obtain three distinct sets of parameters, one for each class.

2. Combine these sets of parameters with a 1NN classifier, and classify your 3 classes. Provide confusion matrices and classification accuracy metrics.

3. We suggest you divide the data (80% of the 432 trial repetitions) into a training dataset, the data you use to fit the model for each of the three classes. Use the remaining 20% to test whether your fit and classification are working properly.

Your second task consists of building a second classifier by means of a reservoir computing network (see document attached). Reservoir computing networks consists of a random interconnected node structure, randomly interconnected in a parametric fashion. This inner structure serves the purpose of projecting your temporal series onto a static structure (the node's activity vector). These activity is then related to an output, typically with a linear or logistic regressor, which in our case should yield a vector label that classifies the dataset in each class. Your goal here is:

1. Using the existing python code for reservoir computing networks, build your network and feed each of the three classes to the classifier independently. You can also use a 1NN to classify the different states provided by the node's activity vector.

2. Again provide classification accuracies and confusion matrices.

3. Compare the first and second classifier, which operational differences are worth mentioning?