

Challenge Data

Crack the neural code of the brain

by GNT ENS

currently running

Introduction

Public ranking

Academic ranking

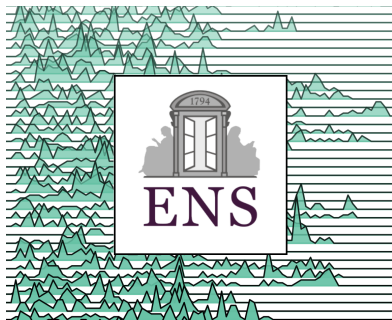
Intermediate academic ranking

Final ranking

Submissions

[Submit a solution](#)

Description



Dates

From Jan. 1, 2019 to Jan. 1, 2020

Challenge context

Neurons in the brain communicate information by generating sequences of electrical pulses (binary events, called spikes). These spiking (binary) sequences from populations of neurons are known to carry information about the current brain activity state of an animal. However, spiking sequences of individual neurons, acting as a neural code, should reflect information processing in the brain as well. Understanding this code and predicting the behavioral state or an action from the underlying neuronal activity is one of the central problems of quantitative neuroscience. Being able to use single neuron spiking data to infer brain state can drastically reduce the number of experiments biological scientists need to perform. At a more fundamental level, the challenge will help quantify how single-cell neural code reflects the brain state.

Challenge goals

The challenge goal is to classify the brain activity state of an animal based on spiking activity patterns of its individual neurons. For this purpose, participants are given

Public metric

cohen_kappa_score from scikit-learn.
[scikit-learn metrics](#)

Course

Not registered with course

[Add a course](#)

Files

x_train

input data of the training set

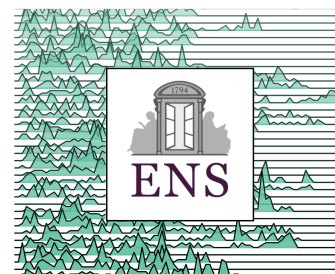
y_train

output data of the training set

x_test

input data of the testing set

The challenge provider



Laboratoire de Neurosciences
Cognitives Computationnel
École normale supérieure

[PROVIDER WEBSITE](#)

recordings of neural spike sequences from the hippocampi of rats. Each spiking sequence in the dataset has a corresponding activity state label (two brain states, labeled STATE1, or STATE2). This is, therefore, a binary classification problem, where each data sample is a time series and participants have to predict which class a given time series sample belongs to.

Join our Slack at bit.do/neuralcode to connect with us and challenge participants and discuss the challenge.

Presentation of the challenge at the Collège de France

You can find the presentation of the challenge made at the Collège de France [here](#) (video in French)

Data description

This problem is a supervised classification problem, so participants are given a number of data samples along with corresponding labels.

In input data files, each line is defined by a unique ID and contains a sample of time-series with 50 values each. These time series are sequences of spike occurrence times (arbitrary time units). Every ID is related to a recording of a single neuron in a certain state. Each line of the input data file also contains a cell identification number called "neuron_id".

The first line of this input file is the header, and columns are separated by commas. The first column corresponds to ID: identification number of the line, it is linked to the label ID provided in the output file. The second column correspond to the neuron identification number ("neuron_id"). The following 50 columns correspond to consecutive values of spike occurrence times.

Here is an example of an input file:

```
ID,neuron_id,timestamp_0,timestamp_1,timestamp_2,
1,2717,0.2406,0.6191,0.9372,
2,668,0.5332,0.5410,2.6912,
3,7678,0.3259,0.3863,0.4667,
```

The training output file contains the target label for each ID, where the target is label of the brain state. The first line of this file is the header and columns are separated by commas. The columns correspond respectively to the identification number of the line and the value of the actual brain state for this specific line as shown in the following sample:

```
ID,TARGET
1,0
2,1
3,0
4,1
```

where target equal to 0 corresponds to the "STATE1" state, and target equal to 1 corresponds to the "STATE2" state. The same neuron (with the same "neuron_id") might be in both "STATE1" and "STATE2", depending on the exact time when the spike timing series is taken from the recording.

The metric used in this challenge to designate the winning participant is the Cohen kappa score. The choice of this metric is motivated by the presence of a class distribution imbalance in the dataset.

Benchmark description

The benchmark solution is produced by converting the series of spike timings to an interval series representation by differencing. Then, feature vectors are computed for each interval series using the tsfresh Python library. A random forest classifier is trained on these feature vectors to get the benchmark solution.

[I want to cancel my participation](#)



[about](#) | [team](#) | [terms of use](#) | [legal notice and privacy policy](#)