Methods for Data Driven Modelling. Tutorial 3 S.C.,J. C., R.M., F.Z.
Replay of the neural activity during sleep after a task

Understanding how memories are stored and consolidated in the brain is one of the fundamental goals of neuroscience. It is known that short term memories are stored in the brain area called the hippocampus while long term memories are stored in the brain area called the prefrontal cortex. It is important to better understand how memories are transmitted from the hippocampus to the cortex and what is the role of sleep for memory consolidation and learning processes. To such aim, multi-electrode recording of a few dozens of neurons in the prefrontal cortex of alive rats have been performed by F. Battaglia's group [1, 2]. A single recording session, in which the same neurons are recorded, is divided into three epochs: a *task* epoch in which the rat has to learn, by trials and errors, what is the rewarded arm in a Y-shaped maze (there are four possible rules which are changed by the operator as soon as the rat learns them: go to the left arm, go to the right arm, go where a light is on, go where a light is off); and two *sleep* epochs, one before and the other after the task. The goal of this tutorial is to analyze the recording data using principal component analysis and tools from random matrix theory, in order to detect similarities between the neuronal activity during the task and during the post-task sleep epochs, which where not present in the pre-task sleep epoch.

Problem

During a recording session, L neurons, labeled by $i = 1, \dots, L$, are recorded for a total time T. Following Refs. [1, 2], the spike train has been discretized in $\Delta t = 100\,ms$ time bin. We call $M = T/\Delta t$ the total number of bins, and t_b the time corresponding to the center of the bth bin, $b = 1, \dots, M$. The activity of a neuron i in a time bin $[t_b - \Delta t/2, t_b + \Delta t/2]$ is represented by a spike count variable $s_i(t_b) = 0, 1, 2, \dots$. Note that for small time bin (e.g. $\Delta t = 10\,ms$), this is equivalent to the binary representation.

Pearson correlation matrix: We define the mean and variance of s_i over time as

$$\begin{aligned} p_i &= \langle s_i \rangle = \frac{1}{M} \sum_{b=1}^{M} s_i(t_b) \ , \\ \sigma_i^2 &= \frac{1}{M-1} \sum_{b=1}^{M} \big[s_i(t_b) - p_i \big]^2 \ . \end{aligned} \tag{1}$$

The *z-score* variables, which will be our data, are defined as

$$y_{bi} = \frac{s_i(t_b) - p_i}{\sigma_i} \tag{2}$$

which defines a $M \times L$ matrix Y. Each line $\boldsymbol{y}_i$ of the matrix defines a realization of the spiking process of the L neurons, and we have M such realizations available, each corresponding to a different time. We define the $L \times L$ Pearson correlation matrix as:

$$C = \frac{1}{M} Y^T Y \qquad \text{or} \qquad C_{ij} = \frac{1}{M} \sum_{b=1}^{M} y_{bi} y_{bj} \ . \tag{3}$$

The matrix element C_{ij} encodes the correlation, averaged over time, of a pair of neurons i, j. Such matrix can be diagonalized to find its eigenvectors and eigenvalues. The central part of the spectrum of the eigenvalues is due to sampling noise, and therefore it follows a Marchenko-Pastur law of parameter $r = L/M < 1$:

$$\rho(\lambda) = \frac{1}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{r\lambda} \ , \qquad \lambda_\pm = \left(1 \pm \sqrt{r}\right)^2 \ . \tag{4}$$

In order to extract the signal components, we must select the eigenvalues that fall outside the interval $[\lambda_-, \lambda_+]$.

Replay: We want to identify the similarity in the correlation structure between the *sleep* neuronal activity (in the *pre-task* or *post-task* epochs) and the *task* activity. Following Peyrache et al. [2], the *task* epoch is called the *template* epoch, while the *pre-task* or *post-task* sleep epochs are called *match* epochs. An instantaneous measure of similarity in the correlated activity is obtained by computing the scalar product, over all the pairs of neurons, of the simultaneous co-activation of the pair during the match and the corresponding pair correlation during the template,

$$R^{match}(t_b) = \frac{1}{2} \sum_{i \neq j} y_{bi}^{match} \, C_{ij}^{template} y_{bj}^{match} \;, \tag{5}$$

which, after averaging over the time bins, gives the total replay:

$$R_T^{match} = \frac{1}{2} \sum_{i \neq j} C_{ij}^{match} \, C_{ij}^{template} \approx \frac{1}{2} \mathrm{Tr}[(C^{match} - I) \; (C^{template} - I)] \;, \tag{6}$$

where Tr is the trace operation, I is the identity matrix, and we used that, by definition, $C_{ii} = (M-1)/M \sim 1$. Using the eigenvalue decomposition $C_{ij}^{template} = \sum_k \lambda_k \, v_i^k \, v_j^k$ in Eq. (5), one obtains the time-dependent replay as a sum of replays associated to each principal component:

$$R^{match}(t_b) = \sum_k \lambda_k R_k^{match}(t_b) \;, \tag{7}$$

where

$$R_k^{match}(t_b) = \frac{1}{2} \left[\left(\sum_i y_{bi}^{match} \; v_i^k \right)^2 - \sum_i \, (y_{bi}^{match} \; v_i^k)^2 \right] \;, \tag{8}$$

keeping in mind that $\boldsymbol{v}^k$ are the eigenvectors associated to the template epoch.

Data:

The spiking activity of a population of $L = 37$ prefrontal cortex neurons during the *pre-task* sleep (`Dati_181014_sleep1.txt`), *task* (`Dati_181014_maze.txt`) and *post-task* sleep (`Dati_181014_sleep2.txt`) epochs can be downloaded from the course webpage, see Tutorial 3 repository. Each file contains a two-column array $\{t_i, i\}$, where t_i is a spiking time of neuron i (times are in units of $0.1\,ms$), ranging between T_{min} and T_{max}, and i is the neuron index. Data are taken from Ref. [1].

Questions:

You are given a starting notebook that reads the data and answers to questions 1 and 2. You can use it to start working more quickly. Complete the notebook or write your own code to answer the following questions.

1. Read the data and produce the raster plot for the L neurons and the first 10 seconds of the activity. Note that in reading the data, you should find T_{min} and T_{max} for each epoch.

2. Compute the Pearson correlation matrix of the *task* activity and diagonalize it. Compare the spectrum of eigenvalues of C^{task} to the Marchenko-Pastur distribution to extract the signal components associated to large eigenvalues. How many signal components with $\lambda > \lambda_+$ are present?

3. Reshuffle the data, by randomly permuting the activity of the single neurons over the M possible time bins, in such a way that the correlations between neurons spiking together are lost, but the frequencies of individual neurons are kept constant. Verify that the spectrum of the eigenvalues of the Pearson correlation matrix of the shuffled data is well described by the Marchenko-Pastur distribution.

4. Identify the neurons that contribute the most to the largest component of the *task* activity. These neurons identify a cell assembly of neurons coactivating during the task.

5. Calculate the total replay of the *pre-task* sleep and *post-task* sleep epochs, with respect to the *task* epoch.

6. Plot the replay as a function of time of the first component in the *pre-task* and *post-task* sleep epochs. Is it constant in time? Compare with the results in Ref. [2].

7. Plot the PCA reactivation trajectories on the plane of the first and second PCA component in the task, by projecting the activit,y as a function of time, along the first and second component. Compare the plot obtained by projecting the activity in the task and the one in the Sleep Pre and Sllep Post, always on the PC1 and PC2 in the maze. Redo the plots for the projections in PC1 and PC3 and PC2 and PC3. Compare the results for the trajectories in the task with the findings in Fig. 1 in [1].

References

[1] *Replay of rule-learning related neural patterns in the prefrontal cortex during sleep*, A. Peyrache, K. Benchenane, M. Khamassi, S. I. Wiener, F. P. Battaglia, Nature Neuroscience, DOI:10.1038/nn.2337 (2009).

[2] *Principal component analysis of ensemble recordings reveals cell assemblies at high temporal resolution*, A. Peyrache , K. Benchenane, M. Khamassi, S. I. Wiener, F. P. Battaglia, Journal of Computational Neuroscience, DOI:10.1007/s10827-009-0154-6 (2009).