Signal and Data Analysis in Neuroscience
Bar-Ilan, spring 2023

Due date: 29/03/23, 08:00

Project 01

General Instructions:

1. Remember legends, axis labels (for all axes) and units for all graphs/plots.
2. Please add comments to your code and submit only the codes specified in the grading table.
3. Write codes in python.
4. We encourage you to write your solution in English. (5 points bonus)
5. Submit a softcopy (including source code) to biu.sigproc@gmail.com.
6. Note the due date and time.
7. Individual work - No code sharing.

1) Paper assignment: Tuning Curves and SNR (40 points)

Prof. De-Cervantes performed an experiment to record extracellularly from a neuron in the eye of a Donnis Qixotenis. This neuron reduces its spiking rate when presented with various wavelengths of light. Using this recordings, Prof. De-Cervantes would like to test the theory of Butts and Goldman 2006 (link). In the following question, you will be guided what parts to read of this paper to generate similar results.

About the data:
14 different wavelengths were presented (labeled 1-14). 50 trials were made for each wavelength. A trial consists of 1 second recording at a sampling rate of 1000 samples/s, the stimulus is presented for 0.4 seconds and starts 0.2 second after recording began.
The file 'q1data' contains a 3d array with size 14X50X1000 containing the results (provided in .pkl format).

A. Explain the two possible interpretations of the tuning curve as explained in Butts and Goldman paper. In what conditions is there a contradiction between the two and why? How can one test the two possible interpretations using Prof. De-Cervantes recordings?
   When reading the text, refer to the SSI as a measurement of SNR and not of information.
   Relevant paragraphs in the text:
   - Abstract
   - Introduction
   - First result: first two paragraphs
   - Second result

   5 points

To evaluate the ideas of Butts and Goldman, you will calculate the tuning curve of the recorded neurons in three different groups of trials: trials with best SNR, trials with worse SNR and all the trials together. We will examine if all trials together are more similar to the best SNR group or worse SNR group to evaluate the neuron's preferred wavelength. Follow the next steps of analysis:

B. Baseline calculation for SNR definition:
   1. Define a duration for baseline activity. Explain the duration you chose for the calculation and your estimation. Justify your estimation using a figure demonstrating the distribution of firing rates in all trials recorded (independently of the wavelength presented).
   2. At each wavelength, find the two sets of 20 trials with the best and worse SNR during the baseline duration. Explain what you defined as signal and what as noise.
   3. Create three arrays:

i. Best 20 SNR trials of each wavelength (14X20X1000)
ii. worse 20 SNR trials of each wavelength (14X20X1000)
iii. All trials recorded (14X50X1000)

Demonstrate the differences between the three groups by one figure which represents the differences in baseline activity (mean and std). The baseline should be calculated for all the wavelengths together, and not for each one separately. As a demonstrative figure you can use a histogram, scatter plots, or any other type to demonstrate the differences in baseline activity. Note to write the SNR of each group next to its name either in the legend, title, or axis titles.

10 points

C. To generate the tuning curve, you need to choose what is the relevant time window to calculate the mean firing rate to compare between the different wavelengths. For each of the data sets you've created, use the following time frames to calculate the average firing rates across trials within each wavelength:
   1. A window of 200 ms which begins with stimulus onset.
   2. A window of 30 ms around the peak response.
   3. A window of 50 ms at the beginning of increase in activity.

   For each window calculate the STD between wavelengths and present the STDs for each window in three figures (one for each data set). Use these figures to decide which of the given time windows is the one you want to use for next calculations and explain your decision.

   10 points

D. Write two functions, each computes the neuron most preferred wavelength. Each function averages across trials, and its output is two vectors- one represents the average calculation for each stimulus, and one of the STD within each stimulus across the trials:
   1. Function 1 computes the tuning curve, the firing rate for each stimulus using the time window you decided of.
   2. Function 2 computes the changes in firing rates between two stimuli next to each other. For each two nearby stimuli calculate the differences between the two along all possible pairs of trials. This vector will be used for computing the output of the function.

   Run the two functions for each dataset and plot their output. Evaluate if the differences in SNR correspond to major differences between the interpretations of the neuron preferred wavelength (based on Butts and Goldman paper). Given the full data, do you think that the tuning curve or the difference curve is more informative?

   15 points

Signal and Data Analysis in Neuroscience
Bar-Ilan, spring 2023

Grading Table

Grade component	**Requirements**	**Points**
Figures	Relevant graphs, correspondence to instructions, no missing components, clear visibility	5
Explanations (Sections A,B,C,D)	150-500 words for each section. Use clear and concise explanations, describing in full details your figures and the conclusions from them. No point reducing for grammar and spelling mistakes. Points will be reduced for overly long, vague explanations.	20
Code writing (Submit code only for section D)	Accurate calculations, comments to explain your code, code is running (no bugs). No point reducing for code efficiency.	10
Argumentation	Presenting persuasive and strong arguments for the claims. Accuracy in the details provided.	5

2) Sampling and quantification (15 points):

Two neuroscientists, Prof. Hemingway and Prof. Garcia-Marquez, are studying the field of sleep:

- Prof. Hemingway experiments are carried out with human subjects using EEG with a temporal resolution of 10 ms.
- Prof. Garcia-Marquez studies are conducted with mice in the hippocampus region using extracellular recordings with a temporal resolution of 1 ms.

Both are studying the effects of a new drug on the transition between two phases of sleep, wakefulness (alpha waves with frequency of 12 Hz) and phase 1 (theta waves with frequency of 6 Hz).

A. Simulate and plot an example single trial recording for each of the methods. Assume that the first half of the recording is alpha wave and the second is theta. Use a scatter plot with the corresponding resolution on the time axis. Duration of the single trial can be different between the two examples.

B. One of the students of Prof. Hemingway offered him to use averaging bins to reduce the noise in the data. Plot two scatter plots, one for bin of 4 samples and one for 5 samples. Add their interpolating plots.
   Explain which bin is better for detecting the sleep phase. Add a detailed calculation.

C. Explain how the use of the wrong temporal resolution can lead to a confusion between the two oscillations. Illustrate your explanation by plotting a scatter plot of the ground truth oscillation and interpolating plot for the false-detected oscillation. Write on the legend the frequency values of each of them.

Grading Table

Grade component	**Requirements**	**Points**
Calculations	Accuracy, full details	3
Explanations	One paragraph for each section with Clear and concise explanations. No point reducing for grammar and spelling mistakes. Points will be reduced for overly long, vague explanations.	8
Figures	Correspondence to instructions, no missing components, clear visibility	4

3) Firing rates and convolution (25 points)

The data file 'spike_times.csv' contains spikes times recorded over 10 seconds. Spikes times are in seconds but recorded with sub-millisecond resolution. Use milliseconds bins to create the spike train.

A. Calculate the mean firing rate over the whole period. **2 points**
B. Find the spike firing rate r(t) using non-overlapping windows of length 0.2s. Repeat this for windows of 0.5s, 1s and 3s.
   Plot r(t) for the different windows in the same plot (in different colors).
   4 points
C. Write your own convolution function 'MyConv'. Compare 'MyConv' to the built-in convolution function (python – numpy.convolve).
   Display the result of convolving a simple uniform function [x = ones(30,1)] with a small rectangular window (length = 6 bins). Plot also the result from the built-in function for comparison. Make sure to treat the edges correctly.
   5 points
D. Find the spike firing rate r(t) using a sliding rectangular window of length 0.2s.
   Repeat this for windows of 0.5s, 1s and 3s.
   Plot r(t) for the different windows on the same plot.
   Remember to normalize each window's area to 1.
   4 points
E. Find the spike firing rate r(t) using a sliding Gaussian (window span = 0.8 sec, std =0.25s).
   Repeat with a Gaussian window of span = 0.45 sec and std = 0.1s.
   Plot r(t) for the different windows on the same plot. Plot both windows.
   3 points
F. Write one-two pages to summarize your results. Display the results of sections A,B,D,E together and explain the differences between them (one page). **7 points**

Notes:

1. In case 'MyConv' results are different form the built-in functions, use the built-in functions for D-E. Otherwise, use your function.
2. Data for the assignment is CSV format with a ',' delimiter, you can load it using the np.genfromtxt('file', delimiter=',').

Grading Table

Grade component	**Requirements**	**Points**
Code writing (Submit code only for section C)	Accurate calculations, comments to explain your code, code is running (no bugs). No point reducing for code efficiency.	5
Figures and calculations	Accuracy, no missing components, clear visibility	13
Explanations	No point reducing for grammar and spelling mistakes. Points will be reduced for overly long, vague explanations.	7

4) Stochastic Point Process (20)

The files 'spk1.csv' and 'spk2.csv' contain spike times of stochastic point processes recorded for 50 seconds. Spike times are in seconds.

A. Write a general function which computes the Fano Factor (FF) and Coefficient of variation (CV) for a given vector of spike times.
   The function must include the following definitions:
   input: signal (1D vector of floats)
   size of window for calculations
   output: Fano Factor value (FF)
   Coefficient of variation (CV)
B. Run your function with the two signals using windows of 1 second, 500 ms and 100 ms. Present your results for each window and explain the differences between them.
C. Determine if any of the given signals fits the Poisson model for a neuron. If one doesn't fit the model, offer another model for its behavior (general description, not mathematical one).

Grading Table

Grade component	Requirements	Points
Code writing (Submit code only for section A)	Accurate calculations, comments to explain your code, code is running (no bugs). No point reducing for code efficiency.	10
Explanations (Sections B,C)	A paragraph for each section. Use clear and concise explanations. No point reducing for grammar and spelling mistakes. Points will be reduced for overly long, vague explanations.	10

Good luck!
SDA team.