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

Due date: 11/06/23, 08:00

Project 05

General Instructions:

1. Please add comments to your code and submit only the codes specified in the grading table
2. We encourage you to write your solution in English. (5 points bonus)
3. Individual work - No code sharing.

1) Paper assignment: Information theory uses (60)

Prof. C. Brontë lab studies propagation of epilepsy in the cortex. She would like to test the claims of W. v. Drongelen et al. 2003 (link) using her recordings. In the following question, you will be guided what parts to read of this paper to generate similar results.

A. Explain the research question of Drongelen's paper and why the researchers find information theory as the most suitable method to study this question.
   Relevant paragraphs in the text:
   - Abstract
   - First paragraph of the introduction
   - First paragraph in the method 'Synchrony, recruitment, and reconfiguration'.

B. Calculate analytically (only calculator, excel and wolfram alpha are allowed) the entropies H_t and $mean(H_n^j)$ from the example in figure 3.
   Relevant paragraphs in the text:
   - Under the method 'the method Quantitative analysis' from the paragraph which starts with 'To study network-related…'.

   Important note: You don't need to add the bias term to your calculations.

C. Write a general function which finds best bin for entropy calculation for a group of neurons. Use the range of bins mentioned in the paper (relevant paragraph starts with 'Each of the entropy values…').
   input: matrix of spikes trains when each row represents different neuron.
   output: best minimal bin for the whole neurons
   Notes:
   - Assume the neurons are sampled in 1000 Hz, and the sampling length (number of columns) can be any length.
   - The output should be same bin for all of them, and not for each one separately

D. Generate simulated data for the function you wrote using 3 Poisson Neurons and explain why tthe output of the function is reasonable. For example, you can generate neurons which their firing rate alters between two possible rates every X ms.

E. The file "spk_mat.npy" contains spike trains for 8 neurons over 72 seconds sampled at 1000 Hz.
   a. Using the function you wrote, find the ideal bin for all the neurons.
   b. Divide each recording of neurons to 9 Seconds trial. Find the effective entropy for each neuron as explained in fig. 3 you read. Note, no need to add the bias term.

c. Using the effective entropies you calculated, calculate the conditional entropy measure or the mutual information measure in order to find for the neurons in rows 3,7 & 8 which other neuron can best predict their activity.

Note: For each of the sections C,D,E write 1-2 paragraph explanation about the algorithm and the line of thought you used for every calculation. No need to attach section E code!

Grading Table

Grade component	**Requirements**	**Points**
Explanations	Clear and concise writing. Full answers. No point reducing for grammar and spelling mistakes.	20
Code writing (Submit code only for sections C and D)	Accurate calculations, comments to explain your code, code is running (no bugs). No point reducing for code efficiency.	30 (15 for each section)
Analytical calculations (Section B)	Accuracy, full details, clear explanations.	10

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

2) PCA and Clustering (40)

The data file 'dataPCA.mat' contains a matrix called 'lfp'. Each row of the matrix represents LFP recording of one trial, using a sampling rate of 4KHz.

A. Use a built-in function for PCA and plot the coefficients for the first 3 components.
B. Create a general implementation of PCA (your own). You may use built in functions to calculate eigenvalues and eigenvectors.
   Plot the coefficients for the first 3 components and compare your results to those of the built-in function.
C. For each of the 3 first component:
   - Plot histograms of the projections of each trial onto each component.
   - Plot a scatter-plot of the projections on each pair of 2 components plain.
   - Use built-in functions of K-means and EM to cluster the scatter plot in each graph. Try them for 2-3 different reasonable numbers of clusters.
   - Plot graphs representing each cluster number and clustering method.
     Organize your graphs into 1-3 figures and write a full figure legend for each one.
   - Write an explanation comparing the results of each clustering method. Furthermore, explain the graphs referring to classification and representation of the data.
D. Explain what is the percentage of the variance of the data that is explained by the 1st component? By the 2nd component? By both components combined?
   What is the percentage of the variance that remains unexplained by the 2 components?
E. Find the column (dimension) with the highest variance before PCA and calculate its entropy. Calculate the entropy of data on the first eigen vector (first PC). Explain your results.

Grading Table

Grade component	**Requirements**	**Points**
Code writing (Submit code only for section B)	Accurate calculations, comments to explain your code, code is running (no bugs). No point reducing for code efficiency.	10
Explanations (1-2 paragraph for each section)	Clear and concise line of thought. Full answers. No point reducing for grammar and spelling mistakes.	15
Calculations	Accuracy, full details, clear explanations.	10
Figures	correspondence to instructions, no missing components, clear visibility	5

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

Good luck!
SDA team.