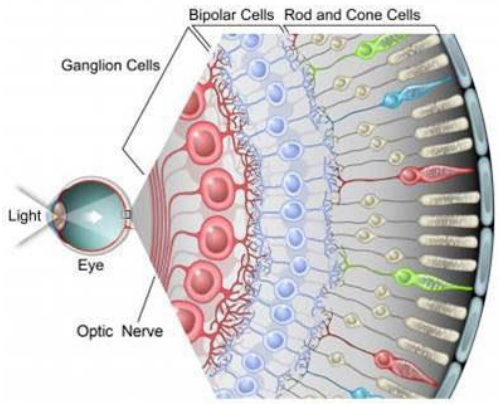
Homework assignment #8

Intermediate report due Wed 11/07; final due in class Mon 11/19

Distinguishing ganglion cell subtypes in neural recordings? (20 points)



Jamie Simon, Salk Institute for Biological Studies.

The schematic on the left shows the structure of the retina. Somewhat counterintuitively, the processing on this picture goes from right to left. The photoreceptors (rod and cone cells) are on the right. This layer is like a CCD of a digital camera – cells respond to the light intensity of the corresponding "pixel". But this information is not sent to the brain directly. Instead, it is pre-processed by a layer of bipolar cells ('receiving' from multiple adjacent photoreceptors each), and then by a layer of Retinal Ganglion Cells ('receiving' from multiple bi-polar cells). Unlike the "pixel"-like photoreceptor cells, the activation of ganglion cells corresponds to more complex features such as movement in the visual field, oriented bright-dark edges, etc. These are the cells whose axons collectively form the "optic nerve" that transmits the visual information into the brain.

Recently, a very active topic of research asks whether ganglion cells come in multiple subtypes, and how many. E.g. this paper from July 2018 uses molecular markers to classify 40 (!) of them: *Rheaume et al.: Single cell transcriptome profiling of retinal ganglion cells identifies cellular subtypes. Nature Communications, 2018. https://doi.org/10.1038/s41467-018-05134-3*

But no one quite knows if all these putative "subtypes" behave differently as far as encoding the visual signal is concerned. And settling the "subtype" from neural recording data is proving difficult. We can indeed distinguish at least a few different patterns of behavior, but how many are there, exactly? And are the distinctions discrete, or is it a continuous spectrum?

This homework will give you a taste of this problem. The attached dataset is from an experiment where a salamander retina was "shown" the same 19-second movie, on repeat, 297 times. The dataset is the signal recorded from 160 ganglion cells all located close together, representing the majority of the ganglion cells in that spatial patch. So roughly speaking, these 160 time-traces of spikes represent the information that a patch of the salamander retina (corresponding to some particular area of the visual field) sent to the brain. Can you identify any subtypes based on the way these cells behave? This is real data, and there is no single "right answer" – see how much you can find, and what behaviors you can identify!

Instructions

- Working in pairs is **highly** encouraged. By talking through your ideas with someone, you will get much further and find more interesting features in the data.

- Download the data and start by exploring it in various ways. Let the data talk to you. How reproducible are the responses across the 297 iterations of the movie? Do some cells consistently fire together? Or earlier? Or later? Are some cells' patterns perhaps more stereotypical, and others' more noisy? Do some cells fire in bursts, are more silent, more intermittent? Etc.
    - When you notice or suspect some interesting features, look for **representations** of data that make them more apparent. For instance, for some purposes, it is sensible to average the response of each neuron across the 297 movie iterations – but not when looking at reproducibility among these iterations. For some questions you may want to look at the average period of silence, or the strength of pairwise correlations between neurons, or the frequency spectrum, etc.
    - Look for neat ways to **visualize** these features that would convince someone else. E.g. an appropriately constructed histogram, a scatter plot, a heatmap...
- Use your observations to **generate a hypothesis** about putative behavioral subtypes. Describe the hypotheses in the intermediate report & include your evidence.
- **Design a similarity metric** appropriate for the representation you are using and the features you are interested in. Justify your choice by explicitly stating: (a) to what features of your data you want your metric to be sensitive; and (b) to what features you want it to NOT be sensitive (give an example).
- Would k-means, or PCA, or another technique discussed in class be appropriate to try? If so, see if it confirms the subtype structure you hypothesized.
- **Important: blindly applying clustering or PCA won't work.** Explore / plot the data in various ways to generate hypotheses first, and then try to find targeted ways to confirm, support, or refute them. Here, "targeted" means using the representation and a similarity metric appropriate for the hypothesis.
- Once again: work in groups, ask questions. Submit one document per group.

The data (description provided by the authors)

Multi-electrode array recording from salamander retinal ganglion cells

This data was collected as part of the study [1]. It consists of preprocessed multi-electrode array recording from 160 salamander retinal ganglion cells responding to 297 repeats of a 19 s natural movie. The data is available in two formats: (1) a .mat file containing an array with dimensions "number of repeats" x "number of neurons" x "time in a repeat"; (2) a zipped .txt file containing the same data represented as an array with dimensions "number of neurons" x "number of samples", where the number of samples is equal to the product of the number of repeats and timebins within a repeat. The time dimension is divided into 20 ms time windows, and the array is binary indicating whether a given cell elicited at least one spike in a given time window during a particular repeat. See the reference below for details regarding collection and preprocessing:

[1] Tkačik G, Marre O, Amodei D, Schneidman E, Bialek W, Berry MJ II. Searching for Collective Behavior in a Large Network of Sensory Neurons. PLoS Comput Biol. 2014;10(1):e1003408.

DOI of this Data Collection: 10.15479/AT:ISTA:61

Related publication DOI: doi:10.1371/journal.pcbi.1003408
Related publication URL: http://dx.doi.org/10.1371/journal.pcbi.1003408