Statistical Neuroscience
Gastpar

Ecole Polytechnique Fédérale, Lausanne: Spring 2014
May 2, 2014

Mini-Project 2
Due: May 27, 2014, at 9:15

Note: Some problems are marked *advanced.* Those problems give only a small number of points when compared to the effort required to solve them. Do these problems last, or not at all, depending on your interest.

1 The Data Set: BMI Data from Macaque

Note: This data is courtesy of Prof. J. M. Carmena, University of California, Berkeley. This fact must be acknowledged in your presentation slides and in your report.

The Experiment: A monkey was performing a behavioral task during about 15 minutes (which is 1080353 milliseconds). The task consisted in holding the hand in the center of a switchboard with eight lightbulbs arranged in a circle around the center. When one of the eight lightbulbs lit up, the monkey had to move the hand to the lightbulb, and then back to the center of the board. This is often referred to as a *reaching task.* During the entire 15 minutes, voltage traces were recorded at 64 sites (using a multielectrode array). From these LFP signals (local field potentials) the spike times of individual neurons can be extracted using a spike sorting algorithm.

In this miniproject we are going to use the mean firing rate of each neuron. More specifically, the mean firing rates corresponding to the time intervals when the monkey was reaching towards one of the 8 targets. For each of these reaching directions 26 trials were recorded during the experiment. To spare some time for you with the data handling, we have already extracted the mean firing rates from the raw data and posted the file `rates_all.mat` on Moodle.

Nonetheless, if you are interested, you are welcome to look at the original files yourself as well. You can download the data from: https://documents.epfl.ch/users/g/ga/gastpar/public/matlabBMI.zip A separate PDF file is posted on Moodle, giving you explanations regarding the full data.

2 Firing Rate Entropies and Information Rates

Using the .mat file posted on Moodle, in this problem, we calculate entropies and information rates for 184 neurons.

(a) (15 Points) To estimate entropies, let us use the histograms. To this end, you may use the matlab command `p = hist(vector,N)/length(vector)`. Here, N is the number of bins you want matlab to use (read the matlab help if you want to know more). Using your entropy function from Homework 2, evaluate the firing rate entropies for all 184 neurons, using different numbers of bins (for example $N = 10, 50, 100$). In a single plot, where the x-axis ranges from 1 to 184 (all our neurons), compare the entropy estimates for various N. What do you observe?

(b) (15 Points) Next, we want to estimate pairwise mutual informations. As far as I know, there is no built-in matlab function that does two-dimensional histograms. So, you'll have to do it yourself. But it's okay to keep things simple: Round every firing rate to the closest integer. Then, simply write a double for-loop where for every combination of an integer firing rate for neuron k and an integer firing rate for neuron ℓ, you count how many occurrences you have in the data. Make sure to normalize the

outcome so that it sums to one (and hence, is a valid probability mass function). Then, use your code from Homework 2 to calculate the information. Do this for a few selected neuron pairs

(c) (advanced) (5 Points) Do Part *(b)* for all neuron pairs. Represent your findings in a clever way — can you identify clusters of neurons (i.e., groups of neurons where inside the group, the mutual information is "large" for every pair)? How many such clusters? What about connections between these clusters? A graph representation may be nice here.

(d) (advanced) (5 Points) Finally, let us estimate information rates via a Gaussian approximation. For two Gaussians with means m_1 and m_2, respectively, and with covariance matrix Σ, the mutual information can be calculated (much like what you did in Homework 2) to be

$$I_{Gaussian} = \frac{1}{2}\log_2 \frac{\Sigma_{11}\Sigma_{22}}{\det \Sigma}, \qquad (1)$$

where, as usual, Σ_{ij} is the element in column i, row j, of the matrix Σ. (Prove it, if you like such calculations!) Fitting a Gaussian (separately) to each neuron pair from Parts *(b)* and *(c)* and plugging into the Gaussian information formula, find alternative values for information. Does this change your insights in Part *(c)*?

3 Decoding reaching directions from firing rates

Your are given a .mat file containing the firing rates of all 184 recorded neurons during 208 trials in total. These trials correspond to different targets and your task will be to figure out which one of the 8 targets the money was reaching to during each trial using the firing rates only. In other words, we have a set of unlabeled data samples (firing rate vectors) which we would like to cluster into 8 groups.

One possible approach is to build a statistical model that tries to capture the hidden structure of the data, namely the reaching directions. In cases where there is only one hidden variable taking discrete values a popular choice is to use a Gaussian mixture model (GMM). For finding the parameters of the GMM you will use the EM algorithm (just like in Homework3 Problem2). Tip: since the dimensionality of the data (184) is much larger than the number of trials we have available (26) it is a good idea to use only a subset of the neurons. The selection criterion can be chosen based on various heuristics e.g. neurons with high firing rate, or neurons with large response entropy.

(a) (25 Points) Run the EM algorithm on the firing rate data. Note for the M step: for estimating the covariance matrix of each cluster you have very limited amount of data hence it is worth constraining the covariance matrix to be diagonal. To avoid the algorithm getting stuck in local minima, run your code several times using different initial conditions. Plot the log likelihood of the data as a function of iterations. Compare the final values across different runs.

(b) (15 Points) After estimating the parameters of the GMM you can compute the posterior distribution $P(v|u)$ over the directions v for each data sample u. Using this, classify each firing rate vector into one of the 8 movement directions. Visualize the result by plotting the label of the direction (1-8) as a function of the sample indices.

(c) (10 Points) Try different assumptions for the form of covariance matrix. E.g $\sigma^2 I$, $diag(\sigma_1^2 ... \sigma_n^2)$, same/different covariance matrix for each cluster. Comment on the results. What are the possible advantages/disadvantages of each choice?

(d) (5 Points) For the EM algorithm you have to provide the number of clusters, however, the right number is often unknown. Experiment with different number of clusters.

(e) (5 Points) What are the assumptions that we made by choosing a Gaussian mixture model? How can the choice of model affect your results? Can you think of some other choices?