G80.3042.002 – Fall 2007
Statistical Analysis and Modeling of Neural Data

Homework 2

Due: 31 Oct 2007

Your results should be in the form of a MATLAB file (typically, the filename should have an extension of `.m`). Email your solutions to `eero@cns.nyu.edu` and `bijan@cns.nyu.edu`.

1. **Decoding a pair of neurons**.

   Imagine two neurons, each generating spikes according to a Poisson process. Consider the problem of discriminating between two stimuli (call them A and B). Assume that the two neurons respond to these stimuli with rates $r_1(A) = r_2(A) = 1$, $r_1(B) = 3$, and $r_2(B) = 12$, in units of spikes during the (arbitrary) fixed time interval that the stimulus is presented.

   (a) Compute the ML decision rule neuron 2 (by itself). Specifically, generate a binary vector corresponding to the ML decision for all spike counts over the range $[0, 20]$ (timesaver: use `poisspdf` to compute the Poisson probability). Determine the performance (percent correct) for this decision rule, when applied to an equal number of presentations of the two stimuli. Check this by simulating 1000 responses from the model neuron for each stimulus (you can use the function `poissonrnd`), and computing the percentage of correct answers given by your decision rule. Explain (in words) what you would need to change in this construction if the two stimuli were not equally likely to occur (e.g., if you wanted to optimize percent correct but $s = 0$ occurs twice as often as $s = 1$).

   (b) Now compute the ML rule for the joint responses, $\vec{r} = [r_1 r_2]$, assuming they are independent. Specifically, calculate a binary image containing the decision rule for all pairs of spike counts in the range $[0, 20] \times [0, 20]$. Again, compute both the actual and simulated percentage of correct answers for this decision rule.

   (c) Simulate 1000 joint responses for each stimulus, and use these to compute the Fisher Linear Discriminant (see `http://www.cns.nyu.edu/ eero/NOTES/LeastSquares.pdf` if you don't remember how to do this). For this problem, the discriminant is $\vec{d} = (C_A + C_B)^{-1}(\mu_B - \mu_B)$, where the $\{\mu, C\}$ are means and covariances of the two stimulus-conditioned response distributions. Given the discriminant $\vec{d}$, the decision rule requires that we compare the projected responses $\vec{r} \cdot \vec{d}$ to some threshold. Ordinarily, we'd like to do this to maximize the likelihood of *projected* responses, but this is a messy numerical problem. Instead, select this threshold as the value such that $(1 - c(\vec{r_A} \cdot \vec{d})) = c(\vec{r_B} \cdot \vec{d})$, where $c$is the cumulative distribution (compute by generating a histogram, computing the cumulative with `cumsum`, and solving for the crossing point using `interp1` assuming linear interpolation). Plot your decision boundary (a straight line) on top of a plot of the optimal ML rule from the previous part (a binary image). How do they compare? Compute the percentage of correct answers for this decision rule- how does it compare to that of the previous part?

   (d) Now consider a correlated spiking model. Specifically, imagine that the conditional resopnse of the second neuron is Poisson with rate $r_2(A|N_1) = \lfloor 1 + 2N_1 \rfloor$ and $r_2(B|N_1) = \lfloor 12 + 2N_1 \rfloor$. Note this is reminiscent of (but not precisely the same as) what might arise in the GLM model presented in a few weeks ago in class. Compute the joint likelihood for each of the stimuli (you'll need to compute this using $p(N_1, N_2) = p(N_2|N_1)p(N_1)$). Compute the ML decision rule and compare to previous results. Again, repeat using the Fisher Linear Discriminant. Compare the discriminant to the ML rule, and compare the percent correct for a sample data set.

(e) Finally, assume the neurons are anti-correlated, such that rate $r_2(A|N_1) = \lfloor 1 - 0.75N_1 \rfloor$ and $r_2(B|N_1) = \lfloor 12 - 0.752N_1 \rfloor$. Again compute the ML rule, and the FLD solutions, and compare,both visually and in terms of percent correct.

2. **Decoding LFP activity**.

For this problem, you will work with another example data set recorded from area LIP of an awake, behaving monkey. This data set contains simultaneous recordings of three LFP signals at sites a few hundred microns apart during the delay period as a monkey performs delayed reach-and-saccade movement to a peripheral target.

Data format: The data is organized in one data structure `Data` with the fields `Lfp1`, `Lfp2`, `Lfp3`, containing three Lfp recordings, and `Target` containing the target index for each trial. The data is organized in 500 ms per trials with one trial per row. The signals are sampled at 1kHz. The target indices for each trial go from 1-8 around the circle, where 1 is to the right, 2 is up-right, 3 is up and so on until 8, which is down-right.

Software: You have been given a matlab function to estimate smoothed spectral estimates, `dmtspec.m`. Place this program in the path of your matlab session. To estimate the spectrum of an Lfp signal with W Hz smoothing for each trial use the command `spec1 = dmtspec(Lfp1,[.5,W],1e3);`. This will return an array of spectra one for each trial from DC to 500 Hz. `help dmtspec` will return some documentation. When working with spectral power, always take the `log` before analyzing the signal further. it i.e. work with `log(spec1)`.

(a) Estimate the spectral tuning curve for each Lfp signal across all eight directions for each frequency using a 5 Hz smoothing parameter. What is the preferred target direction of each recording? Does it vary with frequency? How can you determine whether the variations at a given frequency are statistically significant? What can you say about the signal at frequencies below $100Hz$? Above $200Hz$? Above $300Hz$?

(b) Select the trials to one preferred target direction common to all recordings and the opposite target direction. Construct a response vector using activity from all three Lfp recordings. For each frequency, separately apply Fisher's linear discriminant to discriminate trials when the target direction was in the preferred and anti-preferred directions. Use the approach given in Question 1c. What is the performance at each frequency?

(c) Assuming the log LFP power is Gaussian-distributed around the tuning curve, and independent across channels, construct the ML rule using the Log-likelihood ratio, again separately for each frequency. Apply the ML rule you have derived to classify the trials into the two classes. What is the performance at each frequency? Compare with the Fisher Linear Discriminant and comment. Are the assumptions you made reasonable ones? How would you modify your model to address its limitations?

(d) Repeat the above exercise using a 40 Hz smooth parameter. How does the performance change? Comment.