Homework 2

It's GLM time! Remember, a GLM is essentially the same as an LN model discussed in the lecture, just expressed in the language of statistics. There are some code templates in MATLAB to help you get started, but you are welcome to write your solution in any programming language.

Fit a generalized linear model to spiking data. The model is:

$$p(y_i|\theta) = \frac{(\lambda_i \Delta)^{y_i} e^{-\lambda_i \Delta}}{y_i!}; \quad \lambda_i = f(\vec{k} \cdot \vec{x}_i + b)$$

where y_i and $\vec{x}_i$ are the number of spikes and the stimulus at time step i, Δ is the time step size, $\vec{k}$ is the stimulus filter, and $f(b)$ is the background firing rate in the absence of a stimulus.

In the files `Problem_train.mat` and `Problem_test.mat`, you will find 4 variables. `Spikes_train` and `Spikes_test` are row vectors where the entries in the vector are the number of spikes in each time step. `X_train` and `X_test` are matrices of dimension $49 \times 200,000$ where 49 is the number of stimulus dimensions and 200,000 is the number of time steps (200 s with Δ of 0.001 s). Fit the model with 3 nonlinearities: the exponential ($f(u) = \exp u$), the smooth rectified linear ($f(u) = \ln(1+\exp u)$), and the rectified quadratic ($f(u) = [u]_+^2$). Note that for the rectified quadratic, add a small value ε so that $\ln f(u)$ is never $-\infty$ (e.g., $f(u) = [u]_+^2 + \varepsilon$ with $\varepsilon = 10^{-5}$). Then calculate the log-likelihood of each of these three fits on the test data. When you train with some data, and test the fit with other data, it is called cross validation, and this allows us to identify the best model among the fits (i.e., the model with the highest cross validated log-likelihood). Once you have identified the the best model, plot $\vec{k}$ for this model using the following MATLAB command `imagesc(reshape(k,7,7))`.

You may wish to use the following MATLAB template files for this problem: `fitGLM_template.m` and `GLM_LL_template.m`. The former is a template for performing Newton–Raphson optimization to get the maximum likelihood solution for $\vec{k}$. The latter is a template for calculating the log-likelihood given the data and parameters, and can thus be used to get the cross validated log-likelihoods.

The submittables for this problem are your code, the three cross-validated log-likelihoods for the test data, and the plot of $\vec{k}$ and value of the background firing rate $f(b)$ for the best model.

If you have done this right, there's a happy ending to this exercise.