Quantitative Methods in Neuroscience
(NEU 466M)
Homework 8
Due: Friday April 12 by 5 pm in NHB 3.128

This homework explores the ability of PCA to do denoising. We also study some commonly encountered probability distributions in Neuroscience. *General guidelines:* Read through each complete problem carefully before attempting any parts. Feel free to collaborate in groups of size 2-3, but always note the names of your collaborators on your submitted homework. For graphs: clearly label your axes and use good color and symbol choices. Print out your matlab code (in the form of a script file). For derivations you're asked to do 'by hand' (in other words, analytically, using paper and pencil) feel free to turn in handwritten or typed-out work.

1) **Denoising with PCA.** Use the file `HandwrittenDigits.mat` from the previous assignment. The `.mat` file contains a matrix of images of handwritten numbers. Each row of the matrix contains a 28x28 pixel image, with elements resorted into a vector. Add to the matrix a random matrix, using `0.2*randn(`l_1`,`l_2`)`, where l_1 and l_2 are the height and width of the handwritten digits matrix, respectively.

   This is your new data matrix. Re-do problem 2) from the previous assignment, using the noisy new matrix as your data matrix. In part c.), reconstruct all the digits using k principal components, varying k from 1 to 784, and plot the average squared reconstruction error between your reconstructed approximation to the noisy data versus the non-noisy (original data) as a function of k. This will be the sum of the pixel-by-pixel squared difference between your PCA reconstruction and the non-noisy digit, averaged over all handwritten digits. While reconstruction accuracy can only increase with number of PCA components in reconstructing the data on which you did the PCA, here I expect the error to decline but then increase when the comparison is with the original non-noisy data because the bottom PCA components capture the noise. Make these plots and interpret. Make a few plots of the reconstructed digits, the noisy data, and the original non-noisy data, for different k at different points in the reconstruction error curve.

2) Computing the mean and variance of some common probability distributions.

   a. Consider the exponential distribution, a continuous distribution defined as

$$P(x) = (1/L)e^{-x/L}$$

where L is some length-constant. Plot the distribution. Verify that this distribution is normalized. Derive the mean and the variance of the exponential distribution. [Hint: Use the trick of differentiating by a constant, and the observation $\int xe^{\alpha x}dx = \frac{d}{d\alpha}\int e^{\alpha x}dx$.] Derive the signal-to-noise ratio (SNR) of this distribution, defined as the mean over the standard deviation (recall that standard deviation is the square-root of the variance).

b. Consider the Bernoulli distribution for discrete variables with probability p of success. Derive its mean and variance. Plot the variance versus p, and compute the SNR of this distribution as a function of p. Interpret.

c. In class, we saw how to derive the mean of the Poisson distribution. Using a similar approach, compute the variance of the Poisson distribution [hint: try writing $k^2 = k(k-1) + k$]. An important quantity for quantifying spike count statistics is the *coefficient of variance* (CV), defined as the variance over the mean. What is the CV of a Poisson process?

3) Spike train analysis of an entorhinal cortical neuron. Download again the spike response of an entorhinal grid cell from our course webpage.

a. Compute the interspike intervals and then plot a histogram of interspike intervals (hint: use `diff`, `hist`).

b. Slide a window of length T over the data, and count the number of spikes in that bin (hint: use `conv` with a boxcar function of length T and height 1 for this). For this window length, obtain the average number of spike counts in that window, and the variance of the number of spike counts. Repeat this for windows of length $T = 10$ milliseconds to $T = 1$ second, and plot the CV versus T. As we saw above, a Poisson process has a coefficient of variance of 1. When the coefficient of variance exceeds 1, the process is said to be super-Poisson. A point process with a smaller coefficient of variance is called sub-Poisson.