HW2: Information Theory Problem Set 1

42-699B/86-595 Neural Data Analysis

Due at noon Tues 10/2/12.

1. (10 pts.) Prove the chain rule for Information: MI(X,Y;S) = MI(X;S) + MI(Y;S|X). Do not do it from the entropy definition of MI, as was done in the reading; instead, do it from the probability-based definition of MI:

$$MI(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

2. (10 pts.) Show that the entropy, H(Y), of a random variable Y distributed according to a Poisson distribution with mean λ is:

$$H(Y) = \lambda(1 - \log(\lambda)) + e^{-\lambda} \sum_{n=0}^{\infty} \frac{\log(n!)\lambda^n}{n!}.$$

3. (10 pts.) Assume you are recording from a neuron in the motor cortex while a subject performs a task consisting of a sequence of trials. On each trial he reaches for one of 2 targets located in different spatial locations and simultaneously, for each trial, you record the number of times a neuron spikes in a given window of time for each trial.

   Let P(T = t) be the prior distribution for the subject reaching for each target. Assume that P(T = 1) = P(T =2) = .5. Further, assume that you have the following probabilities of observing different counts conditioned on each target:

	P(Y = 0|T)	**P(Y = 1|T)**	**P(Y = 2|T)**	**P(Y = 3|T)**	**P(Y = 4|T)**
T = 1	.4	.3	.2	.1	0
T = 2	0	.1	.2	.3	.4

   3a. (5 pts) What is H(Y)?

   3b. (5 pts) What is MI(T;Y)?

4. (15 pts.) The uniform distribution has maximum entropy.

   4a. (5 pts.) Prove that if a random variable X takes on N distinct values with equal probability, then H(X)=$\log_2$(N).

   4b. (10 pts.) Prove that the uniform distribution has maximum entropy. (Hint: Take the KL-distance between the uniform distribution and any other arbitrary distribution p(x)…)

5. **(50 pts.) Matlab simulation: Bias in MI estimates.**

Mutual information can be hard to measure when you don't have much data, especially when you need to empirically estimate the underlying probability distributions. Typically, data insufficiency causes the MI to be over-estimated. You will demonstrate this process in simulation.

You are recording from a neuron in primary visual cortex while presenting gratings at 4 different orientations (0°, 45°, 90°, and 135°). Assume the neuron you are recording from responds to each stimulus with a spike count that is drawn from a Poisson distribution with rate parameters λ_0, λ_{45}, λ_{90}, and λ_{135}, respectively. Let S be the random variable associated with the stimuli, and R be the random variable associated with the response.

5a. (2pts) If the stimuli are presented an equal number of times, what is H(S)?

5b. (2pts) Suppose λ_0=2, λ_{45}=4, λ_{90}=6, and λ_{135}=8. Compute H(R|S=0°), H(R|S=45°), H(R|S=90°), and H(R|S=135°), to the nearest hundredth of a bit. (Hint: you could use the formula in Eqn. 2. The Matlab function gammaln can be very useful for evaluating log(n!). An alternate possibility is to compute p(r|s) using the function poisspdf.)

5c. (2 pts) Compute H(R|S).

5d. (2pts) Compute H(R).

5e. (2pts) Compute MI(R;S).

5f. (10 pts) Assume each stimulus is presented N times. Simulate the responses you might record by drawing N times from a Poisson distribution for each stimulus (use poissrnd). Now estimate p(r|s) empirically, by counting the number of times you observed each spike count when each particular stimulus was played, and dividing by N. With these probabilities, you can compute MI(R;S). Plot MI(R;S) as a function of N, for N=[4,8,16,32,63,128]. Include in the figure a plot of the real MI from (4d) as a red dotted line. Be sure to label your axes!

5g. (10 pts) Repeat (5f) 100 times for each value of N. Plot the mean and standard deviation of MI(S;R) as a function of N. (Use the Matlab function errorbar.) Be sure to include the real MI as a red dotted line.

5h. (5 pts) Describe in general terms the results you see in 5g.

5i. (10 pts) Now, instead of estimating the probabilities empirically, instead estimate λ_0, λ_{45}, λ_{90}, and λ_{135} from your data, and use these estimates to compute MI, assuming the neurons are Poisson distributed. Repeat 100 times for each value of N, and use errorbar to plot the mean and std of the estimated MI as a function of N. Again, include the real MI as a red dotted line.

5j. (5 pts) Describe the results in 5i. Why wouldn't you always estimate MI this way?

6. (5 pts.) About how much time did you spend on each question of this problem set? Which problem taught you the most, and which taught you the least?