Assignment 7
Theoretical Neuroscience

Peter Dayan, Maneesh Sahani

November 26, 2011

1. **Perceptron**

   (a) Construct a perceptron (equation 8.46) that classifies 10 binary inputs according to whether their sum $\sum u_a$ is positive or negative. Use a random set of binary inputs during training and compare the performance (both the learning rate and the final accuracy) of the Hebbian (equation 8.47), delta, and perceptron learning rules.

   (b) Repeat this training protocol, but this time attempt to make the output of the perceptron classify according to the parity of the inputs, which is the sign of their product $\prod u_a$.

   (c) Why is this example so much harder than the first case?

2. **Indirect and direct actor**

   (a) Implement a stochastic three-armed bandit using the indirect actor (equation 9.14) and the action choice softmax rule (equation 9.12). Let arm a produce a reward of 1 with probabilty p_a, with $p_1 = 1/4, p_2 = 1/2, p_3 = 3/4$, and use a learning rate of $\epsilon = 0.01, 0.1, 0.5$ and $\beta = 1, 10, 100$. Consider what happens if after every 250 trials, the arms swap their reward probabilities at random.

   (b) Averaging over a long run, explore to see which values of ϵ and β lead to the greatest cumulative reward. Can you account for this behavior?

   (c) For the case of just two of the bandits, implement the direct actor from the lecture notes (this is actually a different direct actor from the one in the textbook) and compare performance with the indirect actor.

3. **Maximum entropy point processes**

   Consider a point process $\mathcal{P}_\lambda$ with a constant mean rate constrained to be λ. We are interested in the form of the maximum entropy process consistent with this constraint.

   (a) First, consider the stochastic process defined by taking successive inter-event intervals generated by $\mathcal{P}_\lambda$. How does the constraint on $\mathcal{P}_\lambda$'s rate constrain the ISI process? What is the maximum entropy ISI distribution [recall the discussion of maximum entropy distributions from lecture]? What does this imply about $\mathcal{P}_\lambda$?

   (b) Now consider the stochastic process defined by counting events from $\mathcal{P}_\lambda$ that fall in successive intervals of length Δ. What does the mean rate constraint for the point process mean as a constraint for this discretised counting process? What is the maximum entropy counting process under this constraint? Is this consistent with the form of $\mathcal{P}_\lambda$ you obtained above?

(c) Suppose we were to expect spike trains in the brain to achieve maximum entropy with constrained spike rate. Which of the two preceding approaches to the obtaining the maximum entropy distribution is likely to be the more relevant to the brain. [Hint: how does the process obtained in the second case depend on Δ? Is there a preferred Δ for the brain?]

4. **Communication through a probabilistic synapse**

(a) The Blahut-Arimoto algorithm.

In this part of the question, we derive an algorithm to find an input distribution that achieves the capacity of an arbitrary discrete channel.

i. Given a channel characterised by the conditional distribution $P(R|S)$, we wish to find a source distribution $P(S)$ that maximises the mutual information $I(R;S)$. Show that

$$I(R;S) \geq \sum_{s,r} P(s)P(r|s) \log \frac{Q(s|r)}{P(s)}$$

for any conditional distribution $Q(S|R)$. When is equality achieved?

ii. Use this result to derive (in closed form) an iterative algorithm to find the optimal $P(S)$.* This is called the Blahut-Arimoto algorithm. Prove that the algorithm converges to a unique maximal value of $I(R;S)$.

* Hint: by analogy to EM, alternate maximisations of the bound on the right hand side with respect to Q and to $P(S)$.

(b) Synaptic failure.

Many synapses in the brain appear to be unreliable; that is, they release neurotransmitter stochastically in response to incoming spikes. Here, we will build an extremely crude model of communication under these conditions.

Assume that the input to the synapse is represented by the number of spikes arriving in a 10 ms interval, while the output is the number of times a vesicle is released in the same period. Let the minimum inter-spike interval be 1 ms (taking into account both the length of the spike and the refractory period), and assume that at most 1 vesicle is released per spike. Thus, both input and output symbols on this channel are integers between 0 and 10 inclusive.

Let the probability of vesicle release be independent for each spike in the input symbol, and be given by α^n where α is a measure of synaptic depression and n is the number of spikes in the symbol. (We are neglecting order-dependent effects within each 10ms symbol, and any interactions between successive symbols. This is a terrible model of synaptic behaviour).

i. Generate (in MATLAB) the conditional distribution of output given input for this synapse. Take $\alpha = 0.9$. Use Blahut-Arimoto to derive the capacity-achieving input distribution and plot it.

ii. Try to interpret your result intuitively. Might this have anything to do with the short "bursts" of action potentials found in many spike trains?

iii. `OPTIONAL`: Improve on the model of synaptic transmission. Consider 5 ms input and output symbols, each being a 5-bit binary number where a 1 indicates a spike or a vesicle release. The probability of transmission for each spike in the symbol is again α^n but n is now the number of vesicles released *so far* for this symbol. Construct a new conditional distribution table and repeat the optimisation. Do you get a qualitatively similar result?