CO6 "Introduction to Computational Neuroscience"

Lecturer:
Boris Gutkin
Ecole normale supérieure
29 Rue d'Ulm, GNT, 2nd Floor, right
Email:boris.gutkin@ens.fr

Tutor:
Gregory Dumont
Ecole normale supérieure
29 Rue d'Ulm, GNT, 2nd Floor, right
Email: gregory.dumont@ens.fr

course website: http://iec-lnc.ens.fr/group-for-neural-theory/teaching-260/article/co6-course

Exercise Sheet 2 — 13 March 2017
Please submit your solution in the next class

(1) **Temporal-difference learning with discounting.** In many instances, immediate rewards are worth more than those in the future. To take this observation into account, the value $V(s_t)$ of a particular state s_t is not the sum of all future rewards, but rather the sum of all future, *discounted* rewards,

$$V(s_t) = r(s_t) + \gamma r(s_{t+1}) + \gamma^2 r(s_{t+2}) + \ldots = \sum_{\tau=0}^{\infty} \gamma^\tau r(s_{t+\tau}) \tag{1}$$

where $0 < \gamma < 1$. Here s_t is the state at time t, i.e., the state in which the agent is right now, s_{t+1} the state that the agent will move to next and so on. Following the derivation in the lecture, show that the temporal-difference-learning rule in this case is given by

$$V(s_t) \rightarrow V(s_t) + \varepsilon(r(s_t) + \gamma V(s_{t+1}) - V(s_t)) \tag{2}$$

(2) **Models for the value function.** In the lecture, we talked about the necessity to introduce models for the value of a state, so that one could properly generalize to new, unseen situations. One very simple model is given by the value function $V(\mathbf{u}) = \mathbf{w} \cdot \mathbf{u}$ where $\mathbf{u}$ is a vector of stimuli that could either be present (1) or absent (0).

(a) Take the example of two stimuli, $\mathbf{u} = (u_1, u_2)$. Let us assume that the subject (agent) has already learned the value of a state in which the first stimulus is present, and the value of a state in which the second stimulus is present. The learned values are given by

$$V(\mathbf{u} = (1, 0)) = \alpha$$
$$V(\mathbf{u} = (0, 1)) = \beta$$

What are the values of the parameters $\mathbf{w} = (w_1, w_2)$ that the agent has learnt? Now we assume that the agent, for the very first time, runs into a state in which both stimuli are present. What is the value of this state? What if we now add some uncertainty. What would be the value of a state where the first stimulus has 50% chance of being present and the second stimulus has 10%? In what situation do you think this sort of a more generalized model that you just came up with would not make much sense?

(b) **Advanced:** Derive the temporal-difference learning rule for the parameters $\mathbf{w}$ that need to be learned if the value function is $V(\mathbf{u}) = \mathbf{w} \cdot \mathbf{u}$. Hint: Start from a loss function - what would be a suitable choice? Can you also derive a learning rule if the value function were given by $V(\mathbf{u}) = f(\mathbf{w} \cdot \mathbf{u})$ with $f(\cdot)$ being a known (non-linear) function?