15-883 Computational Models of Neural Systems

Homework 7

This assignment gives you some practice with the Temporal Difference (TD) learning rule. Review the equations in Lecture 5.1.

Assume we have a single stimulus x_1 that comes on at the start of a trial (time t=0) for one time unit. Assume a reward of 100 is supplied at time t=3.

Recall that V(t) is the total discounted future reward:

$$V(t) = \sum_{\tau=0}^{\infty} \gamma^{\tau} r(t+\tau+1) = r(t+1) + \gamma V(t+1)$$

1. Assume a discount factor γ of 0.9. Assume that r(t) and V(t) are zero if t>4. Fill in the values for just V(t) in the table below. Note: you will need to work backward from V(4).

t	$x_1(t)$	r(t)	V(t)	δ(t)
0	**1**	0		
1	0	0		
2	0	0		
3	0	**100**		
4	0	0		

2. Recall that the reward prediction error δ(t) is the difference between the predicted future reward at time t, V(t), and the actual reward at time t+1 plus the discounted expected future reward γV(t+1):

$$\delta(t) = r(t+1) + \gamma V(t+1) - V(t)$$

Since we have a fixed ISI and the reward does not vary, there is no uncertainty, so V(t) can be a perfect predictor and there should be no prediction error. Use the formula to fill in δ(t) in the table above by showing each of the three terms in δ(t), e.g., write something like 0 + 0.9×0 − 0 = 0.

[Continued on next page.]

3. The animal cannot know V(t), so it must learn an estimate V*(t). Let's use a single linear neuron to calculate V*(t). Assume a complete serial compound representation of x_1, so we have values $x_{1,0}$ through $x_{1,4}$. We also have corresponding weights $w_{1,0}$ through $w_{1,4}$. We estimate V*(t) by:

$$V^*(t) = \sum_{i=0}^{4} w_{1,i} \cdot x_{1,i}(t)$$

Since the stimulus comes on at time t=0, the buffer value $x_{1,i}(t)$ is 1 when i=t and 0 otherwise. Therefore $V^*(t)$ is just $w_{1,t}$. The prediction error is calculated using V*(t) since V(t) isn't known, so:

$$\delta(t) = r(t+1) + \gamma V^*(t+1) - V^*(t) \qquad with\ \gamma = 0.9$$

Assume that all the weights start out at 0, and the learning rate is η=0.1. The TD learning rule is:

$$\Delta w_{1,i}(t) = \eta \cdot \delta(t) \cdot x_{1,i}(t) \qquad \text{and} \qquad w_{1,i}(t+1) \leftarrow w_{1,i}(t) + \Delta w_{1,i}(t)$$

Every weight is updated at every time t, but since $x_{1,i}$ is zero unless i=t, only $w_{1,t}$ actually changes. Using the learning rule above, you can calculate how the weights will look at the end of each trial. Fill in the tables below. Use the current row of the Weights table to calculate the next row of the Predictions table. Then use that row of the Predictions table to calculate the next row of the Errors table. Then, use that row of the Errors table to calculate the next row of the Weights table. Repeat two more times.

Weights	**$w_{1,0}$**	**$w_{1,1}$**	**$w_{1,2}$**	**$w_{1,3}$**	**$w_{1,4}$**
Initial	0	0	0	0	0
After 1 trial					
After 2 trials					
After 3 trials		2.52			

Predictions	**V*(0)**	**V*(1)**	**V*(2)**	**V*(3)**	**V*(4)**
Trial 1					
Trial 2					
Trial 3			19		

Rewards	**r(0)**	**r(1)**	**r(2)**	**r(3)**	**r(4)**
	0	0	0	100	0

Errors	**δ(0)**	**δ(1)**	**δ(2)**	**δ(3)**	**δ(4)**
Trial 1					
Trial 2					
Trial 3					