Fall 2025

CS5368 Intelligent Systems

Assignment 3 Problem solving

First Name	
Last Name	
Student ID	
Due date	November 17th (before the class for 001 and by the end of the day for D01)
Max grade	40

Please answer the following questions and submit them through Canvas. Be sure to submit it to the Assignment 3 problem-solving link.

Problem 1 [20 pts]: Model-Based, TD, and Direct Evaluation RL

An agent interacts with an environment with three states: S1, S2, and S3 (the terminal state). It can take two actions: a1 and a2. During exploration, the agent observes the following transitions:

From	Action	To	Reward	Count
S1	a1	S2	2	7
S1	a1	S1	0	3
S1	a2	S2	2	4
S1	a2	S3	5	6
S2	a1	S1	1	5
S2	a1	S3	5	5
S2	a2	S2	0	8
S2	a2	S3	5	2

Use the following parameters to answer your questions: (1) Discount factor: $\gamma = 0.9$; (2) Learning rate: $\alpha = 0.5$. (3) Initial values: V(S1) = V(S2) = 0, V(S3) = 0 (terminal state)

a. [6 pts] Using the observed counts, compute the estimated model by computing T and R

s	a	s’	T(s,a,s’)	R(s,a,s’)
S1	a1	S2		
S1	a1	S1		
S1	a2	S2		
S1	a2	S3		
S2	a1	S1		
S2	a1	S3		
S2	a2	S2		
S2	a2	S3		
S2	A2	S1		

b. [4 pts] Using the sequence of transitions as an episode:

[(S1, a1, S2, 2), (S2, a2, S3, 5)]

Compute the reward and estimated value for each state visited (S1 and S2) using Direct Evaluation.

c. [10 pts] Using the following sequence of observed transitions, represented by (s,a,s',r), perform temporal difference updates for V(S1) and V(S2)

[(S1, a1, S2, 2), (S1, a2, S3, 5), (S2, a1, S1, 1), (S2, a2, S3, 5)]

Problem 2 [20 pts]: Feature-Based Representation

Consider the following feature-based representation of the Q-function

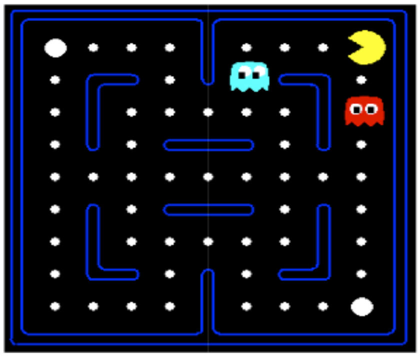
$$Q(s,a) = w_1 f_1(s,a) + w_2 f_2(s,a)$$

with

$$f_1(s,a) = 1/(\text{Manhattan distance to nearest dot after having executed action } a \text{ in state } s)$$

$$f_2(s,a) = (\text{Manhattan distance to nearest ghost after having executed action } a \text{ in state } s)$$

a. [8 pts] Initially, assume $\boldsymbol{w_1 = 1, w_2 = 10.}$ For the state $\boldsymbol{s}$ shown below, find the following quantities. Assume that the red and blue ghosts are both sitting on top of a dot.

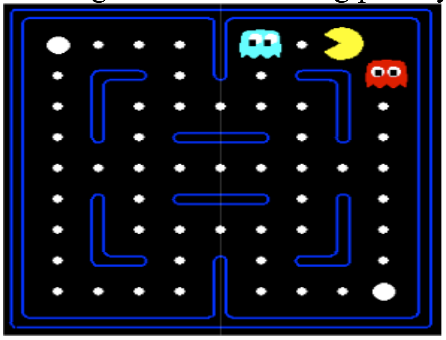


[3 pts] $Q(s, South) =$

[3 pts] $Q(s, West) =$

[2 pts] Based on this approximate Q-function, which action would be chosen? Justify

b. [6 pts] Assume Pac-Man moves West. This results in the state ***s'*** shown below. Pac-Man receives reward 9 (10 for eating a dot and -1 living penalty).



[2 pts] $Q(s', East) =$

[2 pts] $Q(s', West) =$

[2 pts] What is the sample value (assuming γ=1)?

$$sample\ = [r + \ \gamma \max_{a'} Q(s', a')] =$$

c. [6 pts] Now let's compute the update to the weights. Let α=0.5.

[2 pts] $difference\ = [r + \gamma\ \max_{a'} Q(s', a')] - Q(s, a) =$

[2 pts] $w_2 \leftarrow w_2 + \alpha(difference) f_2(s, a) =$

[2 pts] $w_1 \leftarrow w_1 + \alpha(difference) f_1(s, a) =$