0/17 Questions Answered

HW 11 (Electronic Component)

STUDENT NAME

Search students by name or email... ▾

Q1 Q-Learning Properties

5 Points

In general, for Q-Learning to converge to the optimal Q-values...

- □ It is necessary that every state-action pair is visited infinitely often.
- □ It is necessary that the learning rate α (weight given to new samples) is decreased to 0 over time.
- □ It is necessary that the discount γ is less than 0.5.
- □ It is necessary that actions get chosen according to $\arg\max_a Q(s, a)$.

Save Answer

Q2 Exploration and Exploitation

12 Points

Q2.1

2 Points

For each of the following action-selection methods, indicate which option describes it best.

A: With probability p, select $argmax_a Q(s,a)$. With probability $1-p$, select a random action. $p = 0.99$

○ Mostly exploration

○ Mostly exploitation

○ Mix of both

Save Answer

Q2.2

2 Points

Following Part 1:

B: Select action a with probability $P(a \mid s) = \frac{e^{Q(s,a)/\tau}}{\sum_{a'} e^{Q(s,a')/\tau}}$ where τ is a temperature parameter that is decreased over time.

○ Mostly exploration

○ Mostly exploitation

○ Mix of both

Save Answer

Q2.3

2 Points

Following Part 1:

C: Always select a random action.

○ Mostly exploration

○ Mostly exploitation

○ Mix of both

Save Answer

Q2.4

2 Points

Following Part 1:

D: Keep track of a count, $K_{s,a}$, for each state-action tuple, (s,a), of the number of times that tuple has been seen and select $argmax_a[Q(s,a) - K_{s,a}]$.

○ Mostly exploration

○ Mostly exploitation

○ Mix of both

Save Answer

Q2.5

4 Points

Which of the above method(s) would be advisable to use when doing Q-Learning?

□ A

□ B

□ C

□ D

Save Answer

Q3 Feature-Based Representation: Actions

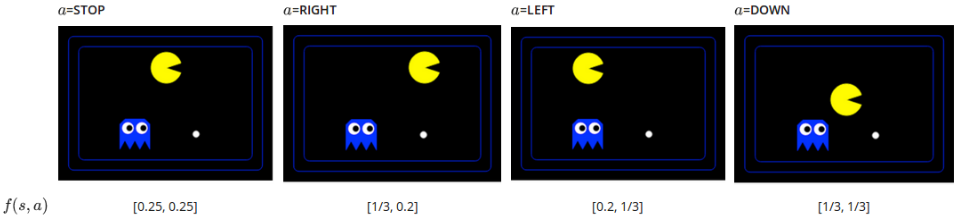
6 Points

A Pacman agent is using a feature-based representation to estimate the $Q(s, a)$ value of taking an action in a state, and the features the agent uses are:

- $f_0 =$ 1/(Manhattan distance to closest food + 1)
- $f_1 =$ 1/(Manhattan distance to closest ghost + 1)

The images below show the result of taking actions STOP, RIGHT, LEFT, and DOWN from a state A. The feature vectors for each action are shown below each image.

For example, the feature representation $f(s = A, a = \mathbf{STOP}) = [1/4, 1/4]$.



The agent picks the action according to $\arg\max_a Q(s, a) = w^T f(s, a) = w_0 f_0(s, a) + w_1 f_1(s, a)$, where the features $f_i(s, a)$ are as defined above, and w is a weight vector.

Q3.1

3 Points

Using the weight vector $w = [0.2, 0.5]$, which action, of the ones shown above, would the agent take from state A?

○ STOP

○ RIGHT

○ LEFT

○ DOWN

Save Answer

Q3.2

3 Points

Using the weight vector $w = [0.2, -1]$, which action, of the ones shown above, would the agent take from state A?

○ STOP

○ RIGHT

○ LEFT

○ DOWN

Save Answer

Q4 Feature-Based Representation: Update

18 Points

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

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

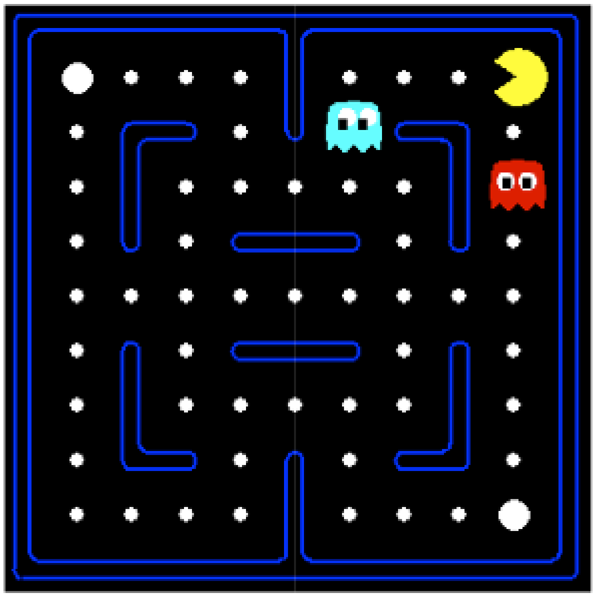
$$f_2(s,a) =$$
$$(\text{Manhattan distance to nearest ghost after having executed action } a \text{ in}$$

Q4.1

2 Points

Assume $w_1 = 1, w_2 = 10$.

For the state s shown below, find the following quantities. Assume that the red and blue ghosts are both sitting on top of a dot.



$Q(s, West) =$

Enter your answer here

Save Answer

Q4.2

2 Points

Following Part 1, $Q(s, South) =$

Enter your answer here

Save Answer

Q4.3

2 Points

Following Part 2, based on this approximate Q-function, which action would be chosen:

○ West

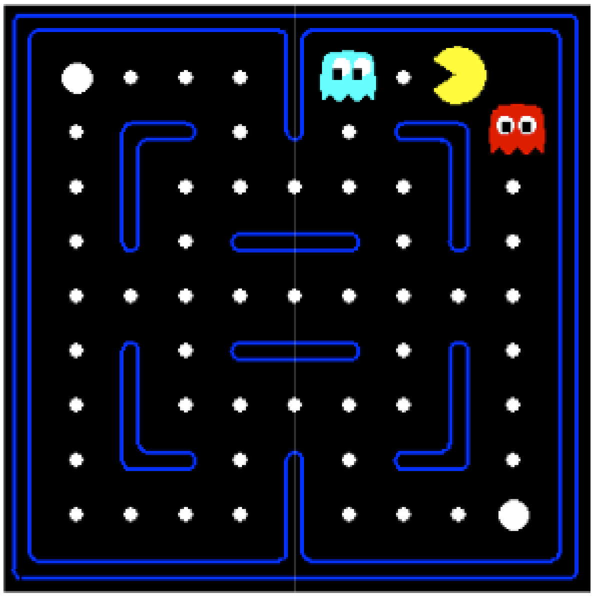
○ South

Save Answer

Q4.4

2 Points

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).



$Q(s', West) =$

Enter your answer here

Save Answer

Q4.5

2 Points

Following Part 4, $Q(s', East) =$

Enter your answer here

Save Answer

Q4.6

2 Points

Following Part 5, what is the sample value (assuming $\gamma = 1$)?

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

Enter your answer here

Save Answer

Q4.7

2 Points

Now let's compute the update to the weights. Let $\alpha = 0.5$.

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

Enter your answer here

Save Answer

Q4.8

2 Points

Following Part 7, $w_1 \leftarrow w_1 + \alpha \, (\text{difference}) \, f_1(s, a) =$

Enter your answer here

Save Answer

Q4.9

2 Points

Following Part 8, $w_2 \leftarrow w_2 + \alpha\,(\text{difference})\, f_2(s,a) =$

Enter your answer here

Save Answer

Save All Answers

Submit & View Submission ›