

hw5_rl_q9_feature_based_representation_update

Question 9: Feature-Based Representation: Update

0.0/18.0 points (graded)

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

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

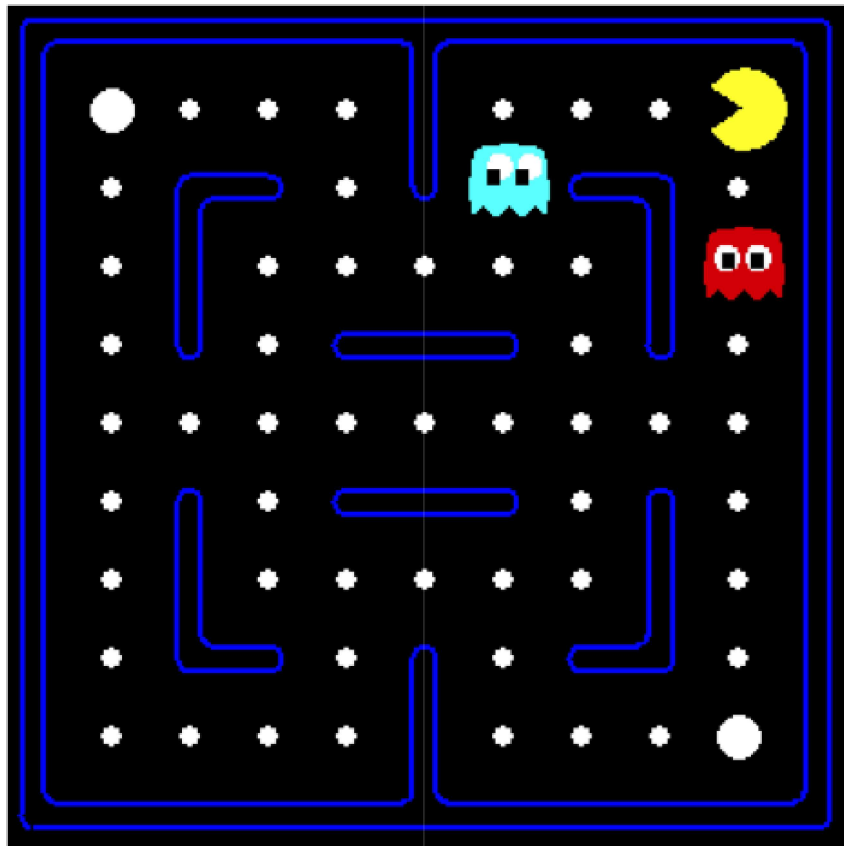
with

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

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

Part 1

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.



$Q(s, West) =$

31

Answer: 31

$$Q(s, West) = 1 * 1 + 10 * 3 = 31$$

$Q(s, South) =$

11

Answer: 11

$$Q(s, South) = 1 * 1 + 10 * 1 = 11$$

Based on this approximate Q-function, which action would be chosen:

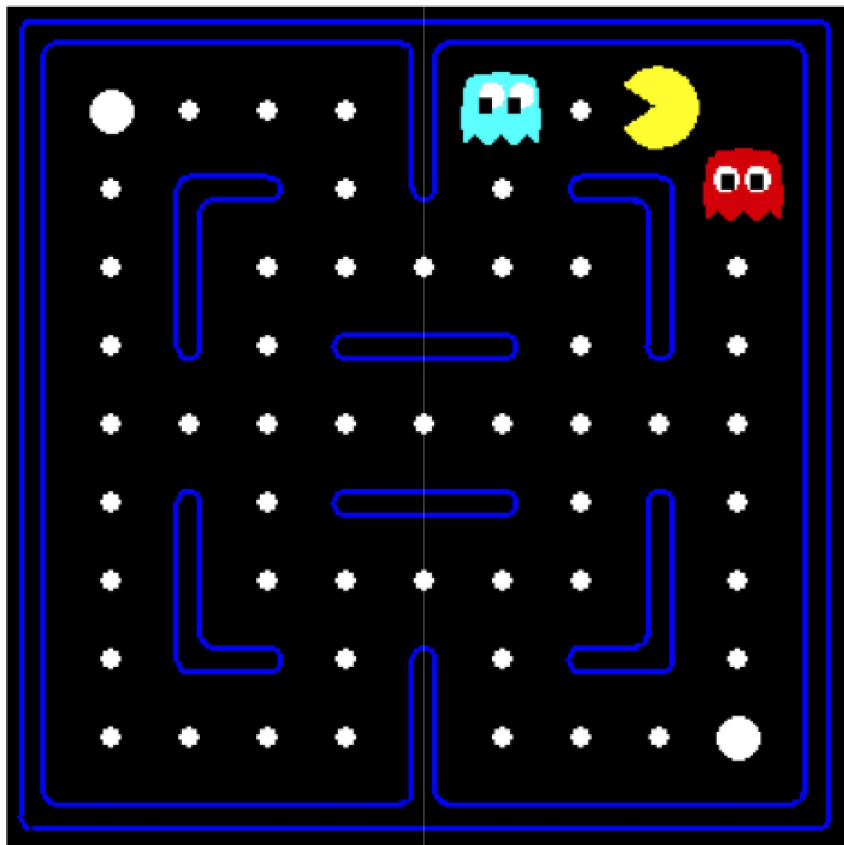
☒ West ✓

☐ South

31 > 10.5, so West would be chosen

Part 2

Assume Pac-Man moves West. This results in the state s' shown below.



The reward for this transition is $r = +10 - 1 = 9$ (+10: for food pellet eating, -1 for time passed).
Fill in the following quantities. Assume that the red and blue ghosts are both sitting on top of a dot.
 $Q(s', West) =$

Answer: 11

$$Q(s', West) = 1 * 1 + 10 * 1 = 11$$

$$Q(s', East) =$$

Answer: 11

$$Q(s', East) = 1 * 1 + 10 * 1 = 11$$

What is the sample value (assuming $\gamma = 1$)?

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

Answer: 20

$$\text{sample} = 9 + 1 * 11 = 20$$

Part 3

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

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

Answer: -11

$$\text{difference} = 20 - 31 = -11$$

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

Answer: -4.5

$$w_1 = 1 + .5 * (-11) * 1 = -4.5$$

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

Answer: -6.5

$$w_2 = 10 + .5 * (-11) * 3 = -6.5$$

For this problem, you may press "Check" as many times as you want without resetting the problem, so that you don't have to reset the problem for trivial math mistakes.