Practice Quiz: Q-function approximation and reward shaping (not assessed) Results for Xinyu Zeng

Submitted Jun 6 at 12:54

Jnanswered

Question 1

0 / 2 pts

Consider an MDP with four actions (A, B, C) and three variables:

x : boolean
y : boolean
z : [1, 10]

How many cells would we have in a Q-table?

○ 4

orrect Answer

○ 160

○ 12

○ 120

○ 32

○ 3

For variable x and y, each has two possible values, while for variable z, it has 10 possible values. Therefore, the number of states is 2 x 2 x 10=40.

The number of actions is four.

Therefore, the number of cells in the Q-table is 160.

Jnanswered

Question 2

0 / 1 pts

Consider again the above MDP, in which we select two features to represent the state and still have four actions.

How many elements will be in our weight vector w?

○ 4: One for each action

○ 0: Weights are independent of features.

orrect Answer

○ 8: One for each state-action pair

○ 2: One for each feature

○ 6: One for each action plus one for each state

Our feature vector has an entry for every state-action pair. If there are two features and four actions, the total number of elements of the state-action feature vector is 8.

Each element in the state-action feature vector has a weight, so the answer is 8.

Jnanswered

Question 3

0 / 2 pts

Consider the Freeway example from lectures. If we have just four features:

1) r: row number relative to the final row, which is normalised to [0,1]
2) dc: distance to the nearest car in the current row, which is normalised to [0,1]
3) da: distance to the nearest car in the row *above,* which is normalised to

[0,1]
4) db: distance to the nearest car in the row *below*, which is normalised to [0,1]

Consider we have just two actions: Up and Down.

If we have the weight vector w = (0.4, 0.3, 0.2, 0.01,   0.2, 0.2, 0.01, 0.2), in which the first four elements are for action Up, and the rest for action Down.

Given state *s* that is row 4, one car in each row that are 2 columns away, what will *f(s, Down)* return if we assume the order of actions in the element is *Up* and then *Down*?

○ (0.4, 0.2, 0.2, 0.2, 0, 0, 0, 0)

orrect Answer ○ (0, 0, 0, 0, 0.4, 0.2, 0.2, 0.2)

○ (0.4, 0.2, 0.2, 0.2)

○ (0.4, 0.4, 0.2, 0.2, 0.2, 0.2, 0.2, 0.2)

○ (0.4, 0.2, 0.2, 0.2, 0.4, 0.2, 0.2, 0.2)

The vector will have 8 elements: one for each state-action pair. The entire vector is returned.

Down is the 2nd set of actions in the vector, and all other parts of the vector (which are the values for action Down), are set to zero because they are not relevant to the Down action; so the answer is (0, 0, 0, 0, 0.4, 0.2, 0.2, 0.2)

The vector is ordered by action, so the answer starting with (0.4, 0.4, 0.2, ...) is not correct this interleaves the values for the action.

Jnanswered

Question 4

0 / 2 pts

Consider the above example of Freeway with four features.

If we have the weight vector w = (0.4, 0.3, 0.2, 0.01,   0.2, 0.2, 0.01, 0.2), in which the first four elements are for action Up, and the rest for action Down, and we are in the state s that is row 4, one car in each row that are 2 columns away, what is Q(s, Down)?

Assume that rows and columns are both min-max normalised 10 rows and 10 columns; so if we are in row 6, the feature value would be 6/10 = 0.6.

ou Answered

orrect Answers

0.162 (with margin: 0)

First, we need to calculate f(s, Down).

The normalised row value is 4/10 = 0.4, and for the other three, the normalised cars values are each 2/10 = 0.2.

Thus f(s, Down) = (0, 0, 0, 0, 0.4, 0.2, 0.2, 0.2)

To calculate Q(s, Down) we take the product of w and f(s, Down):

Q(s, Down) = 0.4*0 + 0.3*0 + 0.2*0 + 0.01*0 + 0.2*0.4 + 0.2*0.2 + 0.01*0.2 + 0.2*02.

= 0 + 0 + 0 + 0 + 0.08 + + 0.04 + 0.002 + + 0.04

= 0.162

Jnanswered

Question 5

0 / 4 pts

Consider the above example of Freeway with four features.

If we have the weight vector w = (0.4, 0.3, 0.2, 0.01,   0.2, 0.2, 0.01, 0.2), in which the first four elements are for action Up, and the rest for action Down, and we are in the state s that is row 6, one car in each row that are 2 columns away.

Assume that rows and columns are both min-max normalised 10 rows and 10 columns; so if we are in row 6, the feature value would be 6/10 = 0.6, and that Q(s, Down) = 0.162.

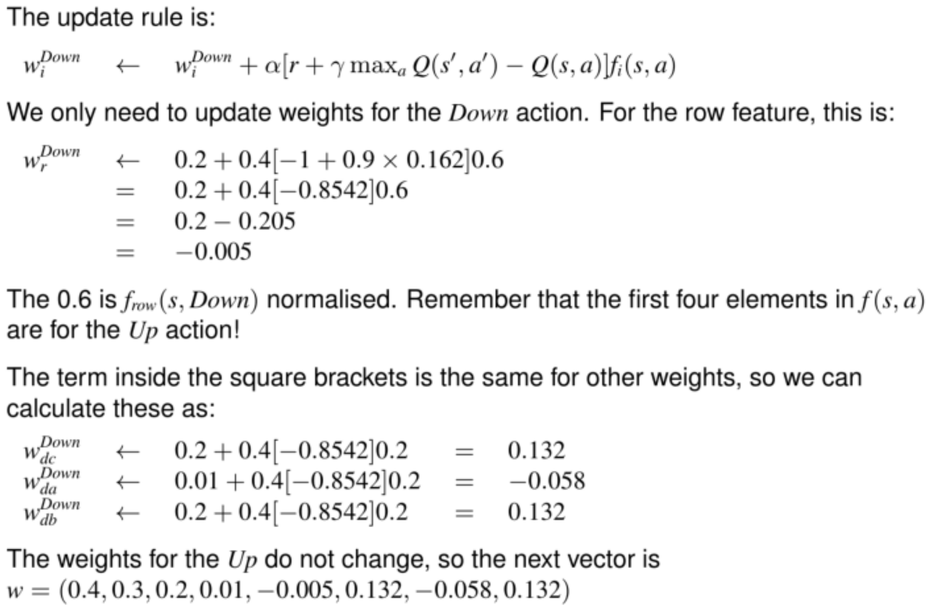
If $\alpha = 0.4$ and $\gamma = 0.9$, and the action Down is executed (going to row 5), and receives a reward of -1. What is the new weight vector using a Q-learning update, assuming that max_a' Q(s',a') = 0.162 and Q(s,a) = 0? Round to three decimal places.

○ w = (-0.005, 0.132, -0.058, 0.132)

○ w = (0.4, 0.3, 0.2, 0.01, 0.005, 0.132, 0.058, 0.132)

○ w = (-0.005, 0.132, -0.058, 0.132, 0.4, 0.3, 0.2, 0.01)

orrect Answer ○ w = (0.4, 0.3, 0.2, 0.01, -0.005, 0.132, -0.058, 0.132)



The update rule is:

$$w_i^{Down} \leftarrow w_i^{Down} + \alpha[r + \gamma \max_{a} Q(s', a') - Q(s, a)]f_i(s, a)$$

We only need to update weights for the *Down* action. For the row feature, this is:

$$\begin{aligned} w_r^{Down} &\leftarrow 0.2 + 0.4[-1 + 0.9 \times 0.162]0.6 \\ &= 0.2 + 0.4[-0.8542]0.6 \\ &= 0.2 - 0.205 \\ &= -0.005 \end{aligned}$$

The 0.6 is $f_{row}(s, Down)$ normalised. Remember that the first four elements in $f(s, a)$ are for the *Up* action!

The term inside the square brackets is the same for other weights, so we can calculate these as:

$$\begin{aligned} w_{dc}^{Down} &\leftarrow 0.2 + 0.4[-0.8542]0.2 &= 0.132 \\ w_{da}^{Down} &\leftarrow 0.01 + 0.4[-0.8542]0.2 &= -0.058 \\ w_{db}^{Down} &\leftarrow 0.2 + 0.4[-0.8542]0.2 &= 0.132 \end{aligned}$$

The weights for the *Up* do not change, so the next vector is $w = (0.4, 0.3, 0.2, 0.01, -0.005, 0.132, -0.058, 0.132)$

Jnanswered

Question 6

0 / 1 pts

Consider the GridWorld example from the notes.

Using the inverse Manhattan distance as a potential reward function, calculate Q(s, West) for state s = (1,2) and state s' = (0,2), receiving no immediate reward.

Assume $\alpha = 0.5$ and $\gamma = 0.9$ and Q(s,a)=0 for all states and actions.

ou Answered

orrect Answers

-0.1 (with margin: 0)

The shaped reward is $F\left(s, s'\right) = \gamma\Phi\left(s'\right) - \Phi\left(s\right)$

If s = (1,2) and s'=(0,2), then $\Phi\left(s\right) = \frac{1}{2}$ and $\Phi\left(s'\right) = \frac{1}{3}$, and therefore $F(s, s') = 0.9 \times \frac{1}{3} - \frac{1}{2} = -0.2$

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

This is a simple calculation:

$$Q(s, a) = 0 + 0.5[0 - 0.2 + 0.9 \times 0 - 0] = -0.1$$