**20/21** Questions Answered
Saved at 12:49 PM

# HW 3

## **Q1** Numerical Answer Formatting

0 Points

Note on formatting (which applies to all questions in this homework)

Many of the questions in this homework have answers that are decimal numbers. Due to current limitations of Gradescope, your answers must be an exact string match to ours. In order to ensure an exact match, please carefully follow the following formatting for your numerical answers.

- For infinite decimals, we will accept any answer within $\pm 0.01$ of the true value of the fraction.
- Do not include any leading or trailing 0s unless they are necessary to show the location of the decimal
- If the number is an integer, do not include a decimal

**Examples:**

`.1234`

`-.001`

`10.4`

`-10`

`0`

Note: If you use the Python interpreter to do your math, floating point error may lead to inexact decimal numbers. It is probably best to use another calculator, but if you do use Python you may need to adjust its output to get the actual exact answer.

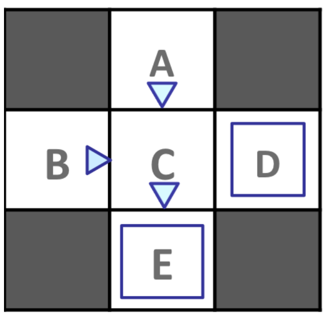
=========

Save Answer

## **Q2** Model-Based RL: Grid

3 Points

**Input Policy $\pi$**




*Assume:* $\gamma = 1$

**Observed Episodes (Training)**

Episode 1

A, south, C, -1
C, south, E, -1
E, exit, x, +10

Episode 2

B, east, C, -1
C, south, D, -1
D, exit, x, -10

Episode 3

B, east, C, -1
C, south, E, -1
E, exit, x, +10

Episode 4

A, south, C, -1
C, south, E, -1
E, exit, x, +10

What model would be learned from the above observed episodes?

T(A, south, C) =

1

**EXPLANATION**

The action south is taken twice from state A, and both times results in state C. $\frac{2}{2} = 1$

T(B, east, C) =

1

**EXPLANATION**

The action east is taken twice from state B, and both times results in state C. $\frac{2}{2} = 1$

T(C, south, E) =

.75

**EXPLANATION**

The action south is taken four times from state C, and results in state E three times. $\frac{3}{4} = .75$

T(C, south, D) =

.25

**EXPLANATION**

The action south is taken four times from state C, and results in state D one time. $\frac{1}{4} = .25$

Save Answer Last saved on **Jul 11 at 11:06 AM**

## **Q3** Model-Based RL: Cycle

14 Points

*We recommend you work out the solutions to the following questions on a sheet of scratch paper, and then enter your results into the answer boxes.*

Consider an MDP with 3 states, A, B and C; and 2 actions Clockwise and Counterclockwise. We do not know the transition function or the reward function for the MDP, but instead, we are given samples of what an agent experiences when it interacts with the environment (although, we do know that we do not remain in the same state after taking an action). In this problem, we will first estimate the model (the transition function and the reward function), and then use the estimated model to find the optimal actions.

To find the optimal actions, model-based RL proceeds by computing the optimal V or Q value function with respect to the estimated T and R. This could be done with any of value iteration, policy iteration, or Q-value iteration. Last week you already solved some exercises that involved value iteration and policy iteration, so we will go with Q value iteration in this exercise.

Consider the following samples that the agent encountered.

| s | a | s' | r |
|---|---|---|---|
| A | Clockwise | B | 0.0 |
| A | Clockwise | B | 0.0 |
| A | Clockwise | B | 0.0 |
| A | Clockwise | C | -10.0 |
| A | Clockwise | C | -10.0 |
| A | Counterclockwise | C | -8.0 |
| A | Counterclockwise | C | -8.0 |
| A | Counterclockwise | B | 0.0 |
| A | Counterclockwise | B | 0.0 |
| A | Counterclockwise | C | -8.0 |

| s | a | s' | r |
|---|---|---|---|
| B | Clockwise | A | -3.0 |
| B | Clockwise | A | -3.0 |
| B | Clockwise | A | -3.0 |
| B | Clockwise | A | -3.0 |
| B | Clockwise | C | 0.0 |
| B | Counterclockwise | A | -10.0 |
| B | Counterclockwise | A | -10.0 |
| B | Counterclockwise | A | -10.0 |
| B | Counterclockwise | A | -10.0 |
| B | Counterclockwise | C | 0.0 |

| s | a | s' | r |
|---|---|---|---|
| C | Clockwise | A | 0.0 |
| C | Clockwise | B | 6.0 |
| C | Clockwise | B | 6.0 |
| C | Clockwise | A | 0.0 |
| C | Clockwise | A | 0.0 |
| C | Counterclockwise | B | -8.0 |
| C | Counterclockwise | B | -8.0 |
| C | Counterclockwise | B | -8.0 |
| C | Counterclockwise | A | 0.0 |
| C | Counterclockwise | B | -8.0 |

## Q3.1
5 Points

We start by estimating the transition function, T(s,a,s') and reward function R(s,a,s') for this MDP. Fill in the missing values in the following table for T(s,a,s') and R(s,a,s').

Discount Factor, $\gamma$ = 0.5

| s | a | s' | T(s,a,s') | R(s,a,s') |
|---|---|---|---|---|
| A | Clockwise | B | M | N |
| A | Clockwise | C | O | P |
| A | Counterclockwise | B | 0.400 | 0.000 |
| A | Counterclockwise | C | 0.600 | -8.000 |
| B | Clockwise | A | 0.800 | -3.000 |
| B | Clockwise | C | 0.200 | 0.000 |
| B | Counterclockwise | A | 0.800 | -10.000 |
| B | Counterclockwise | C | 0.200 | 0.000 |
| C | Clockwise | A | 0.600 | 0.000 |
| C | Clockwise | B | 0.400 | 6.000 |
| C | Counterclockwise | A | 0.200 | 0.000 |
| C | Counterclockwise | B | 0.800 | -8.000 |

M

.6

**EXPLANATION**

The clockwise action was taken 5 times from A, and went to B 3 times. $3/5 = .6$

N

0

**EXPLANATION**

The transition (A,clockwise,B) happened 3 times and had reward 0 every time. $(0 + 0 + 0)/3 = 0$

O

.4

**EXPLANATION**

The clockwise action was taken 5 times from A, and went to C 2 times. $2/5 = .4$

P

-10

**EXPLANATION**

Both of the occurrences of (A,clockwise,C) had reward -10.

Save Answer Last saved on **Jul 11 at 11:13 AM**

## Q3.2

5 Points

Now we will run Q-iteration using the estimated T and R functions. The values of $Q_k(s,a)$, are given in the table below.

| | A | B | C |
|---|---|---|---|
| Clockwise | -4.24 | -3.76 | 0.72 |
| Counterclockwise | -4.56 | -9.36 | -7.76 |

Fill in the values for $Q_{k+1}(s,a)$.

Q(A, clockwise)

-4.984

**EXPLANATION**

$V_k(B) =$
$max(Q_k(B, clockwise), Q_k(B, counterclockwise)) =$
$-3.76$

$V_k(C) =$
$max(Q_k(C, clockwise), Q_k(C, counterclockwise)) =$
$0.72$

$Q(A, clockwise) = T(A, clockwise, B) \times$
$(R(A, clockwise, B) + \gamma V_k(B))+$

$T(A, clockwise, C) \times (R(A, clockwise, C) + \gamma V_k(C))$

$= .6 \times (0 + .5 \times -3.76) + .4 \times (-10 + .5 \times .72) =$
$-4.984$

Q(A, counterclockwise)

-5.336

**EXPLANATION**

$Q(A, counterclockwise) =$
$T(A, counterclockwise, B) \times$
$(R(A, counterclockwise, B) + \gamma V_k(B))+$

$T(A, counterclockwise, C) \times$
$(R(A, counterclockwise, C) + \gamma V_k(C))$

$= .4 \times (0 + .5 \times -3.76) + .6 \times (-8 + .5 \times .72) =$
$-5.336$

Q(B, clockwise)

-4.024

**EXPLANATION**

$V_k(A) =$
$max(Q_k(A, clockwise), Q_k(A, counterclockwise)) =$
$-4.24$

$Q(B, clockwise) = T(B, clockwise, A) \times$
$(R(B, clockwise, A) + \gamma V_k(A)) +$

$T(B, clockwise, C) \times (R(B, clockwise, C) + \gamma V_k(C))$

$= .8 \times (-3 + .5 \times -4.24) + .2 \times (0 + .5 \times .72) =$
$-4.024$

Q(B, counterclockwise)

-9.624

**EXPLANATION**

$Q(B, counterclockwise) =$
$T(B, counterclockwise, A) \times$
$(R(B, counterclockwise, A) + \gamma V_k(A)) +$

$T(B, counterclockwise, C) \times$
$(R(B, counterclockwise, C) + \gamma V_k(C))$

$= .8 \times (-10 + .5 \times -4.24) + .2 \times (0 + .5 \times .72) =$
$-9.624$

Q(C, clockwise)

0.376

**EXPLANATION**

$$Q(C, clockwise) = T(C, clockwise, A) \times (R(C, clockwise, A) + \gamma V_k(A)) +$$

$$T(C, clockwise, B) \times (R(C, clockwise, B) + \gamma V_k(B))$$

$$= .6 \times (0 + .5 \times -4.24) + .4 \times (6 + .5 \times -3.76) = .376$$

Q(C, counterclockwise)

-8.328

**EXPLANATION**

$$Q(C, counterclockwise) = T(C, counterclockwise, A) \times (R(C, counterclockwise, A) + \gamma V_k(A)) +$$

$$T(C, counterclockwise, B) \times (R(C, counterclockwise, B) + \gamma V_k(B))$$

$$= .2 \times (0 + .5 \times -4.24) + .8 \times (-8 + .5 \times -3.76) = -8.328$$

Save Answer Last saved on **Jul 11 at 11:58 AM**

## Q3.3

4 Points

Suppose Q-iteration converges to the following $Q^*$ function, $Q^*(s, a)$.

| | A | B | C |
|---|---|---|---|
| Clockwise | -5.399 | -4.573 | -0.134 |
| Counterclockwise | -5.755 | -10.173 | -8.769 |

What is the optimal action, either Clockwise or Counterclockwise, for each of the states?

A

- (selected) Clockwise
- Counterclockwise

**EXPLANATION**

We select the action with the highest Q value: $-5.399 > -5.755$

.

B

- (selected) Clockwise
- Counterclockwise

**EXPLANATION**

$-4.573 > -10.173$

C

- (selected) Clockwise
- Counterclockwise

**EXPLANATION**

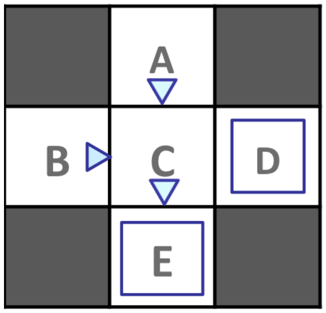
$-0.134 > -8.769$

Save Answer Last saved on **Jul 11 at 11:59 AM**

## **Q4** Direct Evaluation

5 Points

**Input Policy $\pi$**




*Assume:* $\gamma = 1$

**Observed Episodes (Training)**

**Episode 1**

A, south, C, -1
C, south, E, -1
E, exit, x, +10

**Episode 2**

B, east, C, -1
C, south, D, -1
D, exit, x, -10

**Episode 3**

B, east, C, -1
C, south, E, -1
E, exit, x, +10

**Episode 4**

A, south, C, -1
C, south, E, -1
E, exit, x, +10

What are the estimates for the following quantities as obtained by direct evaluation:

$\hat{V}^{\pi}(A) =$

8

$\hat{V}^{\pi}(B) =$

-2

$\hat{V}^{\pi}(C) =$

4

$\hat{V}^{\pi}(D) =$

-10

$\hat{V}^{\pi}(E) =$

10

**EXPLANATION**

The estimated value of $\hat{V}^{\pi}(s)$ is equal to the average value achieved starting from that state.

$\hat{V}^{\pi}(A)$: Episodes 1 and 4 start from state A and both result in a utility of 8. $\frac{8+8}{2} = 8$

$\hat{V}^{\pi}(B)$: Episodes 2 and 3 start from state B. Episode 2 results in -12, while episode 3 results in 8. $\frac{8-12}{2} = -2$

$\hat{V}^{\pi}(C)$: State C is visited in every episode. The remaining rewards from C in episodes 1, 3, and 4 total 9, while the remaining rewards in episode 2 total -11. $\frac{9+9+9-11}{4} = 4$

$\hat{V}^{\pi}(D)$: State D is only visited in episode 2 and has a remaining utility of -10.

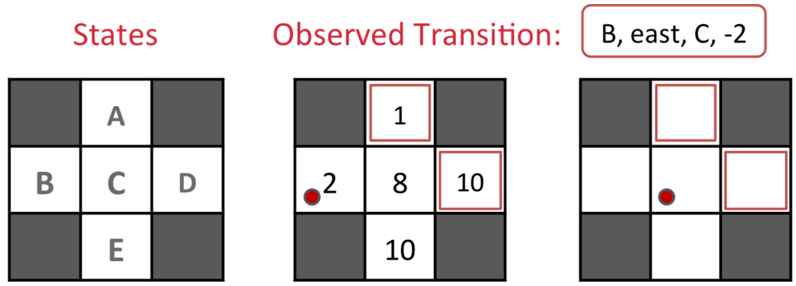
$\hat{V}^{\pi}(E)$: State E is visited in episodes 1, 3, and 4 and has a remaining utility of 10 in each state. $\frac{10+10+10}{3} = 10$

Save Answer Last saved on **Jul 11 at 12:33 PM**

## **Q5** Temporal Difference Learning

5 Points

Consider the gridworld shown below. The left panel shows the name of each state A through E. The middle panel shows the current estimate of the value function $V^{\pi}$ for each state. A transition is observed, that takes the agent from state B through taking action east into state C, and the agent receives a reward of -2. Assuming $\gamma = 1, \alpha = \frac{1}{2}$, what are the value estimates after the TD learning update? (note: the value will change for one of the states only)




*Assume:* $\gamma = 1$, $\alpha = 1/2$

$$V^{\pi}(s) \leftarrow (1-\alpha)V^{\pi}(s) + \alpha\left[R(s,\pi(s),s') + \gamma V^{\pi}(s')\right]$$

$\hat{V}^{\pi}(A) =$

1

$\hat{V}^{\pi}(B) =$

4

$\hat{V}^{\pi}(C) =$

8

$\hat{V}^{\pi}(D) =$

10

$\hat{V}^{\pi}(E) =$

10

**EXPLANATION**

The only value that gets updated is $\hat{V}^{\pi}(B)$, because the only transition observed starts in state B.

$\hat{V}^{\pi}(A) = 1$

$\hat{V}^{\pi}(B) = .5 * 2 + .5 * (-2 + 8) = 4$

$\hat{V}^{\pi}(C) = 8$

$\hat{V}^{\pi}(D) = 10$

$\hat{V}^{\pi}(E) = 10$

Save Answer Last saved on **Jul 11 at 12:33 PM**

## **Q6** Model-Free RL: Cycle

6 Points

We recommend you work out the solutions to the following questions on a sheet of scratch paper, and then enter your results into the answer boxes.

Consider an MDP with 3 states, A, B and C; and 2 actions Clockwise and Counterclockwise. We do not know the transition function or the reward function for the MDP, but instead, we are given with samples of what an agent actually experiences when it interacts with the environment (although, we do know that we do not remain in the same state after taking an action). In this problem, instead of first estimating the transition and reward functions, we will directly estimate the Q function using Q-learning.

Assume, the discount factor, $\gamma$ is 0.5 and the step size for Q-learning, $\alpha$ is 0.5.

Our current Q function, $Q(s, a)$, is as follows.

| | A | B | C |
|---|---|---|---|
| Clockwise | 1.501 | -0.451 | 2.73 |
| Counterclockwise | 3.153 | -6.055 | 2.133 |

The agent encounters the following samples.

| s | a | s' | r |
|---|---|---|---|
| A | Counterclockwise | C | 8.0 |
| C | Counterclockwise | A | 0.0 |

Process the samples given above. Below fill in the Q-values after both samples have been accounted for.

Q(A, clockwise)

1.501

Q(A, counterclockwise)

6.259

Q(B, clockwise)

-.451

Q(B, counterclockwise)

-6.055

Q(C, clockwise)

2.73

Q(C, counterclockwise)

1.460625

Save Answer Last saved on **Jul 11 at 12:49 PM**

## **Q7** Q-Learning Properties

2 Points

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

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

**EXPLANATION**

a)

In order to ensure convergence in general for Q learning, this has to be true. In practice, we generally care about the policy, which converges well before the values do, so it is not necessary to run it infinitely often.

b)

In order to ensure convergence in general for Q learning, this has to be true.

c)

The discount factor must be greater than 0 and less than 1, not 0.5.

d)

This would actually do rather poorly, because it is purely exploiting based on the Q-values learned thus far, and not exploring other states to try and find a better policy.

Save Answer Last saved on **Jul 11 at 12:15 PM**

## **Q8** Exploration and Exploitation

9 Points

### Q8.1

6 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. $p = 0.99$

- (selected) Mostly exploration
- Mostly exploitation
- Mix of both

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

- Mostly exploration
- Mostly exploitation
- Mix of both

C: Always select a random action.

- Mostly exploration
- Mostly exploitation
- Mix of both

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 Last saved on **Jul 11 at 12:30 PM**

## Q8.2
3 Points

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

- [ ] A
- [x] B
- [ ] C
- [x] D

**EXPLANATION**

In general, it is best to use methods that mix exploration and exploitation when doing Q-learning.

Save Answer Last saved on **Jul 11 at 12:16 PM**

## **Q9** Feature-Based Representation: Actions

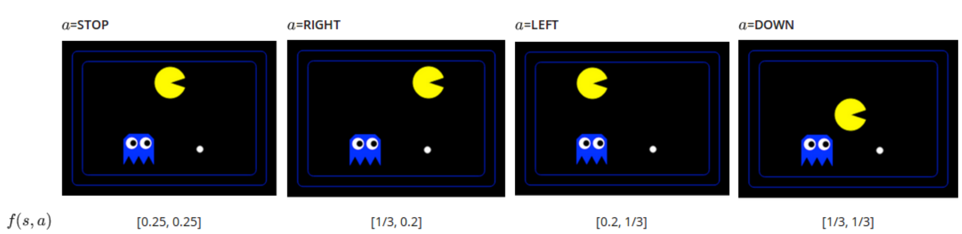
5 Points

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.

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

**EXPLANATION**

STOP:

$$0.2 * 0.25 + 0.5 * 0.25 = 0.175$$

RIGHT:

$$0.2 * 0.33 + 0.5 * 0.2 = 0.166$$

LEFT:

$$0.2 * 0.2 + 0.5 * 0.33 = 0.205$$

DOWN:

$$0.2 * 0.33 + 0.5 * 0.33 = 0.231$$

$0.231$ is the highest value, so the agent would take the DOWN action.

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

**EXPLANATION**

STOP:

$$0.2 * 0.25 - 0.25 = -0.2$$

RIGHT:

$$0.2 * 0.33 - 0.2 = -0.134$$

LEFT:

$$0.2 * 0.2 - 0.33 = -0.29$$

DOWN:

$$0.2 * 0.33 - 0.33 = -0.264$$

$-0.134$ is the highest value, so the agent would take the RIGHT action.

Save Answer Last saved on **Jul 11 at 12:17 PM**

## **Q10** Feature-Based Representation: Update

15 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

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

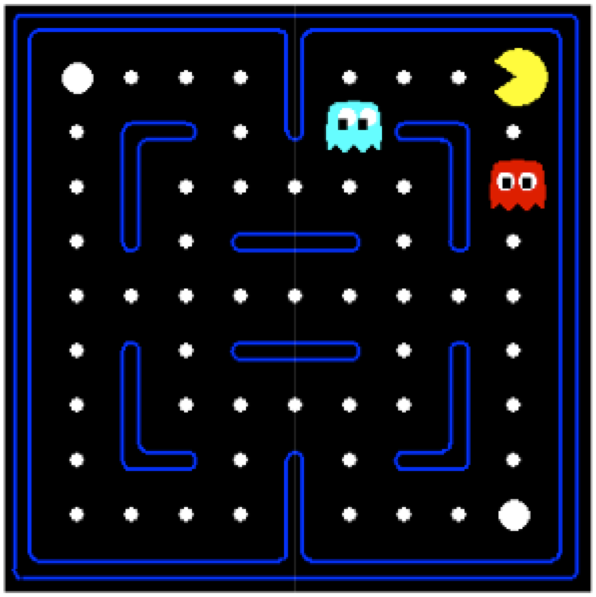
$$f_2(s,a) = (\text{Manhattan distance to nearest ghost after having executed action } a \text{ in}$$

### **Q10.1**

5 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.



$Q(s, West) =$

31

**EXPLANATION**

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

$Q(s, South) =$

11

**EXPLANATION**

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

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

- (●) West
- ( ) South

**EXPLANATION**

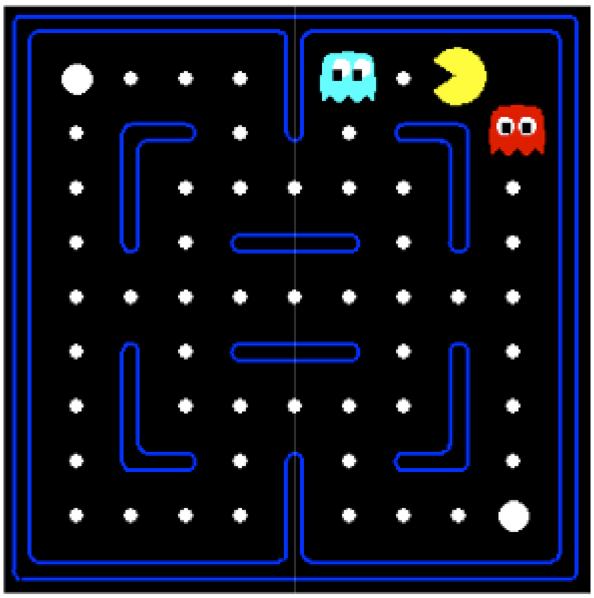
$31 > 11$, so West would be chosen

Save Answer Last saved on **Jul 11 at 12:36 PM**

## Q10.2

5 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).



$Q(s', West) =$

11

**EXPLANATION**

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

$Q(s', East) =$

11

**EXPLANATION**

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

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

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

20

**EXPLANATION**

sample $= 9 + 1 * 11 = 20$

Save Answer Last saved on **Jul 11 at 12:36 PM**

## Q10.3

5 Points

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

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

-11

**EXPLANATION**

difference $= 20 - 31 = -11$

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

-4.5

**EXPLANATION**

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

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

-6.5

**EXPLANATION**

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

Save Answer Last saved on **Jul 11 at 12:36 PM**

## **Q11** Probability, Part I

8 Points

Below is a table listing the probabilities of three binary random variables.
Fill in the correct values for each marginal or conditional probability below.

| $X_0$ | $X_1$ | $X_2$ | $P(X_0, X_1, X_2)$ |
|---|---|---|---|
| 0 | 0 | 0 | 0.160 |
| 1 | 0 | 0 | 0.100 |
| 0 | 1 | 0 | 0.120 |
| 1 | 1 | 0 | 0.040 |
| 0 | 0 | 1 | 0.180 |
| 1 | 0 | 1 | 0.200 |
| 0 | 1 | 1 | 0.120 |
| 1 | 1 | 1 | 0.080 |

### **Q11.1**

4 Points

$$P(X_0 = 1, X_1 = 0, X_2 = 1)$$

0.2

**EXPLANATION**

$P(X_0 = 1, X_1 = 0, X_2 = 1)$: Read from table.

$$P(X_0 = 0, X_1 = 1)$$

0.24

**EXPLANATION**

To find $P(X_0 = 0, X_1 = 1)$, we must compute the probability of all configurations of the variables $X_0, X_1, X_2$ that have $X_0 = 0, X_1 = 1$.

$P(X_0 = 0, X_1 = 1) = P(X_0 = 0, X_1 = 1, X_2 = 0) + P(X_0 = 0, X_1 = 1, X_2 = 1)$

$P(X_2 = 0)$

0.42

**EXPLANATION**

To find $P(X_2 = 0)$, we use the same approach as the last problem. We must compute the probability of all configurations of the variables $X_0, X_1, X_2$ that have $X_2 = 0$.

$P(X_2 = 0) = P(X_0 = 0, X_1 = 0, X_2 = 0) + P(X_0 = 0, X_1 = 1, X_2 = 0)+$

$P(X_0 = 1, X_1 = 0, X_2 = 0) + P(X_0 = 1, X_1 = 1, X_2 = 0)$

Save Answer Last saved on **Jul 11 at 12:36 PM**

## Q11.2

4 Points

$P(X_1 = 0 \mid X_0 = 1)$

.7143

**EXPLANATION**

Use the definition of conditional probability:

$$P(X_1 = 0 \mid X_0 = 1) = \frac{P(X_0=1, X_1=0)}{P(X_0=1)}$$

Calculate $P(X_0 = 1, X_1 = 0)$ and $P(X_0 = 1)$ as in part 2 and 3.

$P(X_0 = 1, X_1 = 0 \mid X_2 = 1)$

.3448

**EXPLANATION**

$$P(X_0 = 1, X_1 = 0 \mid X_2 = 1) = \frac{P(X_0=1, X_1=0, X_2=1)}{P(X_2=1)}$$

$P(X_0 = 1 \mid X_1 = 0, X_2 = 1)$

.5263

**EXPLANATION**

$$P(X_0 = 1 \mid X_1 = 0, X_2 = 1) = \frac{P(X_0=1, X_1=0, X_2=1)}{P(X_1=0, X_2=1)}$$

Save Answer Last saved on **Jul 11 at 12:37 PM**

## **Q12** Probability, Part II

8 Points

You are given the prior distribution $P(X)$, and two conditional distributions $P(Y \mid X)$ and $P(Z \mid Y)$ as below (you are also given the fact that $Z$ is independent from $X$ given $Y$).

All variables are binary variables.

Compute the following joint distributions based on the chain rule.

| $X$ | $P(X)$ |
|---|---|
| 0 | 0.500 |
| 1 | 0.500 |

| $Y$ | $X$ | $P(Y\|X)$ |
|---|---|---|
| 0 | 0 | 0.600 |
| 1 | 0 | 0.400 |
| 0 | 1 | 0.900 |
| 1 | 1 | 0.100 |

| $Z$ | $Y$ | $P(Z\|Y)$ |
|---|---|---|
| 0 | 0 | 0.100 |
| 1 | 0 | 0.900 |
| 0 | 1 | 0.700 |
| 1 | 1 | 0.300 |

## Q12.1

4 Points

$P(X = 0, Y = 0)$

.3

$P(X = 1, Y = 0)$

.45

$P(X = 0, Y = 1)$

.2

$P(X = 1, Y = 1)$

.05

**EXPLANATION**

$P(X, Y) = P(X) \times P(Y \mid X)$

Save Answer Last saved on **Jul 11 at 12:38 PM**

## Q12.2

4 Points

$P(X = 0, Y = 0, Z = 0)$

.03

$P(X=1, Y=1, Z=0)$

.035

$P(X=1, Y=0, Z=1)$

.405

$P(X=1, Y=1, Z=1)$

.015

**EXPLANATION**

In general, from the chain rule, we have that $P(X, Y, Z) = P(X)P(Y \mid X)P(Z \mid X, Y)$. We are given that $Z$ is independent of $X$ given $Y$, hence this simplifies to $P(X, Y, Z) = P(X)P(Y \mid X)P(Z \mid Y)$. We already computed $P(X)P(Y \mid X) = P(X, Y)$ for the previous part, and can re-use those results here.

$$P(X, Y, Z) = P(X) \times P(Y \mid X) \times P(Z \mid Y) = P(X, Y) \times P(Z \mid Y)$$

Save Answer Last saved on **Jul 11 at 12:38 PM**

## **Q13** Probability, Part III

10 Points

For each of the following four subparts, you are given three joint probability distribution tables. For each distribution, please identify if the given independence / conditional independence assumption is true or false.

For your convenience, we have also provided some marginal and conditional probability distribution tables that could assist you in solving this problem.

| $X$ | $Y$ | $P(X,Y)$ |
|---|---|---|
| 0 | 0 | 0.240 |
| 1 | 0 | 0.160 |
| 0 | 1 | 0.360 |
| 1 | 1 | 0.240 |

| $X$ | $P(X)$ |
|---|---|
| 0 | 0.600 |
| 1 | 0.400 |

| $Y$ | $P(Y)$ |
|---|---|
| 0 | 0.400 |
| 1 | 0.600 |

$X$ is independent from $Y$.

- (•) True
- ( ) False

| $X$ | $Y$ | $P(X,Y)$ |
|---|---|---|
| 0 | 0 | 0.540 |
| 1 | 0 | 0.360 |
| 0 | 1 | 0.060 |
| 1 | 1 | 0.040 |

| $X$ | $P(X)$ |
|---|---|
| 0 | 0.600 |
| 1 | 0.400 |

| $X$ | $Y$ | $P(X\|Y)$ |
|---|---|---|
| 0 | 0 | 0.600 |
| 1 | 0 | 0.400 |
| 0 | 1 | 0.600 |
| 1 | 1 | 0.400 |

$X$ is independent from $Y$.

- (•) True
- ( ) False

| $X$ | $Y$ | $Z$ | $P(X,Y,Z)$ |
|---|---|---|---|
| 0 | 0 | 0 | 0.280 |
| 1 | 0 | 0 | 0.070 |
| 0 | 1 | 0 | 0.210 |
| 1 | 1 | 0 | 0.140 |
| 0 | 0 | 1 | 0.060 |
| 1 | 0 | 1 | 0.060 |
| 0 | 1 | 1 | 0.030 |
| 1 | 1 | 1 | 0.150 |

| $X$ | $Z$ | $P(X\|Z)$ |
|---|---|---|
| 0 | 0 | 0.700 |
| 1 | 0 | 0.300 |
| 0 | 1 | 0.300 |
| 1 | 1 | 0.700 |

| $Y$ | $Z$ | $P(Y\|Z)$ |
|---|---|---|
| 0 | 0 | 0.500 |
| 1 | 0 | 0.500 |
| 0 | 1 | 0.400 |
| 1 | 1 | 0.600 |

| $X$ | $Y$ | $Z$ | $P(X,Y\|Z)$ |
|---|---|---|---|
| 0 | 0 | 0 | 0.400 |
| 1 | 0 | 0 | 0.100 |
| 0 | 1 | 0 | 0.300 |
| 1 | 1 | 0 | 0.200 |
| 0 | 0 | 1 | 0.200 |
| 1 | 0 | 1 | 0.200 |
| 0 | 1 | 1 | 0.100 |
| 1 | 1 | 1 | 0.500 |

$X$ is independent from $Y$ given $Z$.

- [ ] True
- [x] False

---

| $X$ | $Y$ | $Z$ | $P(X,Y,Z)$ |
|---|---|---|---|
| 0 | 0 | 0 | 0.140 |
| 1 | 0 | 0 | 0.140 |
| 0 | 1 | 0 | 0.060 |
| 1 | 1 | 0 | 0.060 |
| 0 | 0 | 1 | 0.048 |
| 1 | 0 | 1 | 0.192 |
| 0 | 1 | 1 | 0.072 |
| 1 | 1 | 1 | 0.288 |

| $X$ | $Z$ | $P(X\|Z)$ |
|---|---|---|
| 0 | 0 | 0.500 |
| 1 | 0 | 0.500 |
| 0 | 1 | 0.200 |
| 1 | 1 | 0.800 |

| $Y$ | $Z$ | $P(Y\|Z)$ |
|---|---|---|
| 0 | 0 | 0.700 |
| 1 | 0 | 0.300 |
| 0 | 1 | 0.400 |
| 1 | 1 | 0.600 |

| $X$ | $Y$ | $Z$ | $P(X,Y\|Z)$ |
|---|---|---|---|
| 0 | 0 | 0 | 0.350 |
| 1 | 0 | 0 | 0.350 |
| 0 | 1 | 0 | 0.150 |
| 1 | 1 | 0 | 0.150 |
| 0 | 0 | 1 | 0.080 |
| 1 | 0 | 1 | 0.320 |
| 0 | 1 | 1 | 0.120 |
| 1 | 1 | 1 | 0.480 |

$X$ is independent from $Y$ given $Z$.

- (●) True
- ( ) False

**EXPLANATION**

Part 1 and 2:

Two variables $X, Y$ are independent if $P(X, Y) = P(X)P(Y)$ or equivalently, $P(X|Y) = \frac{P(X,Y)}{P(Y)} = P(X)$. The way to solve this problem is to see if $P(X, Y) = P(X)P(Y)$ or $P(X) = P(X|Y)$ for all combinations of $X, Y$.

Part 3 and 4:

Two variables $X, Y$ are conditionally independent given $Z$ if $P(X, Y|Z) = P(X|Z)P(Y|Z)$ or equivalently, $P(X|Y, Z) = \frac{P(X,Y|Z)}{P(Y|Z)} = P(X|Z)$. You can solve these problems similarly to how you solved the last two problems.

Save Answer Last saved on **Jul 11 at 12:38 PM**

## **Q14** Chain Rule

13 Points

Select all expressions that are equivalent to the specified probability using the given independence assumptions.

Given no independence assumptions, $P(A, B \mid C)$ =

- [ ] $\frac{P(C|A)P(A|B)P(B)}{P(C)}$
- [x] $\frac{P(B,C|A)P(A)}{P(B,C)}$
- [x] $P(A \mid B, C)P(B \mid C)$
- [ ] $\frac{P(A|C)P(B,C)}{P(C)}$

Given that A is independent of B given C, $P(A, B \mid C)$ =

- [x] $\frac{P(C|A)P(A|B)P(B)}{P(C)}$
- [x] $\frac{P(B,C|A)P(A)}{P(B,C)}$
- [x] $P(A \mid B, C)P(B \mid C)$
- [ ] $\frac{P(A|C)P(B,C)}{P(C)}$

Given no independence assumptions, $P(A \mid B, C)$ =

- [ ] $\frac{P(C|A)P(A|B)P(B)}{P(C)}$
- [ ] $\frac{P(B,C|A)P(A)}{P(B,C)}$
- [ ] $\frac{P(A|C)P(C|B)P(B)}{P(B,C)}$
- [x] $\frac{P(C|A,B)P(B|A)P(A)}{P(B|C)P(C)}$

Given that A is independent of B given C, $P(A \mid B, C)$ =

- [x] $\frac{P(C|A)P(A|B)P(B)}{P(C)}$
- [ ] $\frac{P(B,C|A)P(A)}{P(B,C)}$
- [ ] $\frac{P(A|C)P(C|B)P(B)}{P(B,C)}$
- [x] $\frac{P(C|A,B)P(B|A)P(A)}{P(B|C)P(C)}$

Save Answer Last saved on **Jul 11 at 12:40 PM**

Save All Answers

Submit & View Submission >