# ECE421: Introduction to Machine Learning — Fall 2024

## Worksheet 6 Solution: Markov Decision Process

**Q0 (Probability Review)** This question checks your knowledge of basic properties such as independence and conditional independence. Note that the TAs will not review these questions during the tutorial session.

**0.a** Suppose the variables $A$ and $B$ are Boolean variables (*i.e.*, $A$ can have the value $a$ or $\neg a$, and $B$ can have the value $b$ or $\neg b$). Suppose that $A$ is independent of $B$. Determine the missing values in the joint probability mass function $P(A, B)$ below.

| | |
|---|---|
| $P(\neg a, \neg b)$ | 0.1 |
| $P(\neg a, b)$ | 0.3 |
| $P(a, \neg b)$ | |
| $P(a, b)$ | |

**Answer.** $A$ is independent of $B$. Hence, $P(A = x, B = y) = P(A = x)P(B = y)$. Therefore, $\frac{P(\neg a, \neg b)}{P(\neg a, b)} = \frac{\mathbb{P}(B=\neg b)}{\mathbb{P}(B=b)} = \frac{0.1}{0.3} = \frac{1}{3}$. Hence, $\frac{P(a,\neg b)}{P(a,b)} = \frac{\mathbb{P}(B=\neg b)}{\mathbb{P}(B=b)} = \frac{1}{3} \Rightarrow P(a, b) = 3P(a, \neg b)$. These four probabilities should sum to 1. Therefore, $4P(a, \neg b) = 0.6 \Rightarrow P(a, \neg b) = \frac{P(a,b)}{3} = 0.15$. The completed table would be

| | |
|---|---|
| $P(\neg a, \neg b)$ | 0.1 |
| $P(\neg a, b)$ | 0.3 |
| $P(a, \neg b)$ | 0.15 |
| $P(a, b)$ | 0.45 |

**0.b** Suppose $A$, $B$ and $C$ are Boolean variables and that $B$ is independent of $C$ given $A$. Determine the missing values in the joint probability mass function $P(A, B, C)$ below.

| | |
|---|---|
| $P(\neg a, \neg b, \neg c)$ | 0.01 |
| $P(\neg a, \neg b, c)$ | 0.02 |
| $P(\neg a, b, \neg c)$ | 0.03 |
| $P(\neg a, b, c)$ | |
| $P(a, \neg b, \neg c)$ | 0.01 |
| $P(a, \neg b, c)$ | 0.1 |
| $P(a, b, \neg c)$ | |
| $P(a, b, c)$ | |

**Answer.** Observe that $P(A = x, B = y, C = z) = P(B = y \mid A = x, C = z)P(A = x, C = z) = P(B = y \mid A = x)P(A = x, C = z)$, for any $a$, $b$, and $c$. Since $B$ is independent of $C$ given $A$, $P(B = y \mid A = x, C = z) = P(B = y \mid A = x)$, for any $a$, $b$, and $c$. Thus, $P(A = x, B = y, C = z) = P(B = y \mid A = x)P(A = x, C = z)$, for any $a$, $b$, and $c$. Thus, we can rewrite the table as follow.

| | |
|---|---|
| $P(B=\neg b \mid A=\neg a)P(A=\neg a, C=\neg c)$ | 0.01 |
| $P(B=\neg b \mid A=\neg a)P(A=\neg a, C=c)$ | 0.02 |
| $P(B=b \mid A=\neg a)P(A=\neg a, C=\neg c)$ | 0.03 |
| $P(B=b \mid A=\neg a)P(A=\neg a, C=c)$ | |
| $P(B=\neg b \mid A=a)P(A=a, C=\neg c)$ | 0.01 |
| $P(B=\neg b \mid A=a)P(A=a, C=c)$ | 0.1 |
| $P(B=b \mid A=a)P(A=a, C=\neg c)$ | |
| $P(B=b \mid A=a)P(A=a, C=c)$ | |

Rows 1 and 2 imply that $P(B=\neg b \mid A=\neg a)=1/4$, $P(B=b \mid A=\neg a)=3/4$, and $P(A=\neg a, C=\neg c)=1/25$. These results together with row 2 imply that $P(A=\neg a, C=c)=2/25$. Therefore, the value for row 4 is $2/25 \times 3/4 = 0.06$. Moreover, $P(A=\neg a) = 1 - P(A=a) = 3/25$. Rows 5 and 6 and the fact that $P(A=a)=22/25$ imply that $P(A=a, C=c) = 10P(A=a, C=\neg c) = 20/25$. Therefore, $P(B=\neg b \mid A=a)=1/8$ and $P(B=b \mid A=a)=7/8$. Thus, the value for rows7 and 8 are $7/8 \times 2/25$ and $7/8 \times 20/25$, respectively. The completed table would be

| | |
|---|---|
| $P(\neg a, \neg b, \neg c)$ | 0.01 |
| $P(\neg a, \neg b, c)$ | 0.02 |
| $P(\neg a, b, \neg c)$ | 0.03 |
| $P(\neg a, b, c)$ | 0.06 |
| $P(a, \neg b, \neg c)$ | 0.01 |
| $P(a, \neg b, c)$ | 0.1 |
| $P(a, b, \neg c)$ | 0.07 |
| $P(a, b, c)$ | 0.7 |

## Q1 **(Discounting)** Consider the Mars Rover example that we discussed in class. The figure below denote illustrates the environment for this problem, containing five states, namely $a$, $b$, $c$, $d$, and $e$.

| 10 | | | | 1 |
|---|---|---|---|---|
| a | b | c | d | e |

Figure 1: Mars Rover Environment

Assume that the transition dynamics is deterministic. Mars Rover has actions `Left` and `Right` in states $b$, $c$, and $d$, and the only action available in states $a$ and $e$ is `exit`, which ends its mission (*i.e.*, non more transitions after `exit` action). The values in the figure show the reward for the exit action in the two terminal states. Assume a reward of zero for `Left` and `Right` actions.

**1.a** Find the optimal policy when the discount factor is $\gamma = 1$.

**Answer.** It is easy to see that the optimal policy should not cause any infinite loop in Mars Rover's movements. Thus, the best policy should lead to one of the terminal states. It is easy to check that the optimal policy would lead to the terminal state $a$, for all states $b$, $c$, and $d$. With this policy the expected utility in each state would be:

| | a | b | c | d | e |
|---|---|---|---|---|---|
| `Optimal Policy` | `exit` | `Left` | `Left` | `Left` | `exit` |
| `Expected Return` | $10$ | $0 + 1^1 \cdot 10$ | $0 + 1^1 \cdot 0 + 1^2 \cdot 10$ | $0 + 1^1 \cdot 0 + 1^2 \cdot 0 + 1^3 \cdot 10$ | $1$ |

**1.b** Find the optimal policy when the discount factor is $\gamma = 0.1$.

**Answer.**

| | a | b | c | d | e |
|---|---|---|---|---|---|
| `Optimal Policy` | `exit` | `Left` | `Left` | `Right` | `exit` |
| `Expected Return` | $10$ | $0 + 0.1^1 \cdot 10$ | $0 + 0.1^1 \cdot 0 + 0.1^2 \cdot 10$ | $0 + 0.1^1 \cdot 1$ | $1$ |

**1.c** For which $\gamma$ are `Left` and `Right` equally good when in state $d$?

**Answer.** We must have $\gamma^0 \cdot 0 + \gamma^1 \cdot 1 = \gamma^0 \cdot 0 + \gamma^1 \cdot 0 + \gamma^2 \cdot 0 + \gamma^3 \cdot 10$. Thus, $\gamma = \sqrt{1/10}$.

**Q2** **(Multiple Choice)** For each question, choose only one option.

**2.a** Convolutional neural networks have fewer parameters than fully connected networks because they use shared weights. ☑True ☐ False

**2.b** It is best to keep the learning rate of a neural network constant as learning progresses. ☐ True ☑False

**2.c** In a multi-layered neural network, if the activation of a hidden unit is zero, then the gradients of the weights of all of its incoming connections are zero. ☐ True ☑False