תרגיל 3

**Introduction to Artificial Intelligence (236501)**

מגישים: אייל אמדור, בארי זיטלני

318849270, 209351626ת.ז:

חלק א' – MDP ו-RL

חלק א - יבש

שאלה 1

1. הנוסחה עבור התוחלת של התועלת המתקבלת במקרה של ״תגמול על פעולה״ היא:
2. הנוסחה עבור משוואת בלמן:
3. הפונקציה value-iteration:

נשים לב שעבור האלגוריתם לא יתכנס.

1. הפונקציה policy-iteration:

נשים לב שעבור אינסוף מצבים אפשריים ו- *לא מובטח שהאלגוריתם ייתכנס.*

שאלה 2

1. *נגדיר את בעיית ה-*MDP *באופן* *הבא כאשר נתון*:
   1. *קבוצת מצבים* S:
   2. *קבוצת פעולות* A*:*
   3. *מודל מעברים :*

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

* *הערה: מאופן הגדרת מודל המעברים* T *הוא מצב טרמינלי.*
  1. *תגמול*:
  2. *מקדם דעיכה:*

1. *לא ניתן לנסח את הבעיה עם שני מצבים בלבד מאחר וכך לא ניתן יהיה להגדיר בצורה הנדרשת את פונקציית התגמול. בשני מצבים לא ניתן יהיה לדעת איך לאפס את התגמול במידה והקורבן בוחר להתקשר למשטרה (כלומר, כמה שלילי התגמול צריך להיות).*
2. *כן, ניתן לנסח את הבעיה עם ערכים שלילים, לכל מצב* i *זוגי נגדיר תגמול חיובי עם הערך 2, ולכל* i *אי זוגי נגדיר ערך שלילי . באופן הזה בכל פעם שהקורבן שילם סכמנו את התגמול וכאשר הוא מתקשר למשטרה נדע כמה להוריד מסך התגמול כדי לאפס אותו.*
3. *המדיניות האופטימלית מוגדרת באופן הבא:  
   ואילו התועלת האופטימלית מוגדרת כך:  
   כיוון ש הוא מצב טרמינלי מתקיים:  
   על מנת למצוא את נחשב:*

* S=3
* S=2
* S=1
* S=0

*לכן מצאנו כי .  
נמלא את הטבלה כפי שמצאנו.*

|  |  |  |
| --- | --- | --- |
| *תועלות* | *מדיניות* | *ערכי* p |
|  |  |  |
|  |  |  |
|  |  |  |

חלק ב׳ - היכרות עם הקוד

הכרנו.

חלק ג׳ - רטוב

חלק ב' – מבוא ללמידה

שאלה 1

בהינתן

שאלה 2

ייתכן

שאלה 3

אלגוריתמי

שאלה 4

מימוש בקוד.

שאלה 5