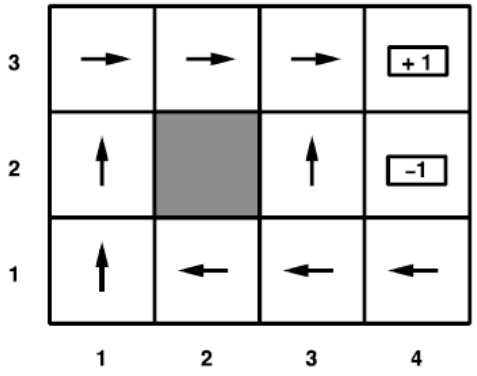
מטלה 2

Grid

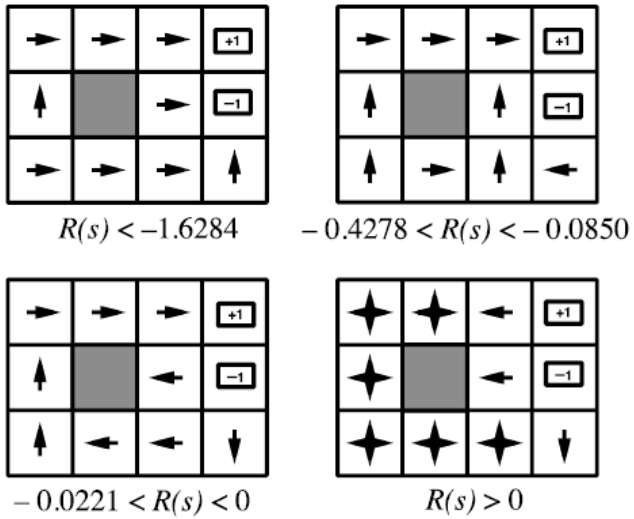
אדמיניסטרציה

- הגשה בזוגות. אין דיון.
- הגשה בתאריך 24.6.
- הגשה במודל.

Grid



(a)



(b)

The optimal policy for R(s) = -0.04

Optimal policies for other choice for R(s)

נתוני הבעיה

- כל משחק מכיל לוח משחק בגודל משתנה הנתון כפרמטר.
- כל צעד על גבי הלוח מתבצע בהסתברות על הצלחת הצעד, גם כן פרמטר לשאלה.
- בלוח המשחק ממוקמים פרסים וקנסות ומיקומם הינו חלק מנתוני הבעיה.
- כל צעד במשחק הינו בעל מחיר, חיובי או שלילי שהינו חלק מתנאי הבעיה.
- נתון discount factor, $\gamma = 0.5$.
- לסיכום, בכדי להגדיר משחק דרושים:
  - גודל הלוח.
  - מיקום הפרסים, עונשים, קירות.
  - הסתברות להצלחת הצעד.
  - עלות כל צעד.

חוקי המשחק

- השחקן יכול לנוע בחופשיות לכל כיוון אבל סיכוי הצלחת התנועה הינם בהתאם לנתוני השאלה.
- אם התנועה לא מצליחה, יכול לנוע ימינה או שמאלה ביחס לתנועה המקורית באופן שווה.
- אם שחקן נתקע בקיר הוא נשאר במקום.
- אם שחקן מגיע למשבצת פרס או קנס המשחק מסתיים.

מטרת המטלה הינה מציאת Policy אופטימלית

- במטלה זו עליכן לממש 3 טכניקות למציאת Policy אופטימלית.

1. אתם נדרשים לממש את משוואות בלמן ולפתור את הבעיה באמצעים אנליטיים ואלגוריתמים שנלמדו בכיתה עד לכדי התכנסות.
2. אתם נדרשים לפתור את הבעיה באמצעות למידת חיזוק RL, ולהגדיר את הבעיה כ - **Model-Based RL**. ניתן לבחור לפתרון כל אלגוריתם הנלמד בכיתה. בנוסף, אתם רשאים לבחור כל שיטה לביצוע Exploration vs. Exploitation.
3. אתם נדרשים לפתור את הבעיה באמצעות למידת חיזוק RL, ולהגדיר את הבעיה כ - **Model-Free RL**. ניתן לבחור לפתרון כל אלגוריתם הנלמד בכיתה. בנוסף, אתם רשאים לבחור כל שיטה לביצוע Exploration vs. Exploitation.

- בכל הסעיפים ניתן להניח שגיאה של לכל היותר $\epsilon = 0.01$ כך שהפרש הערכים בין כל צעד באלגוריתם קטן מ ϵ.

בדיקה

- אני אעביר לכם כ10 בעיות, ימסר בשבוע הבא, 11.6.
- אתם תתבקשו להגיש דוח המשווה בין 3 הגישות לפתרון והתוצאות שהתקבלו. עבור כל בעיה תידרשו להציג את הפתרון שהתקבל המשווה בין ערכים שהתקבלו לכל תא בכל אחת מהגישות. כמו כן, תצטרכו לדווח על הדמיון או השוני בנוגע policy הנבחרה.
- איכות הפתרון תילקח בחשבון.
- חלק מהציון יקבע באופן מדורג על פי התוצאות. לדוגמא: 10 נקודות שיחולקו באופן הבא: התוצאות המובילות יזכו ב10 נקודות והחלשות יזכו בנקודה 1. אז יש מימד חלש של תחרות.

ספציפיקציה במימוש

- תקבלו מימדים לתיאור הW, H grid.
- תקבלו נקודות במרחב לתיאור מיקום פרס (פרס חיובי), עונש (פרס שלילי), קיר (פרס 0) באופן הבא:
  - L=[(x_1,y_1,r_1), (x_2,y_2,r_2),..]
- תקבלו הסתברות על הצלחת צעד p, וכן נדרש לחשב את (p-1)/2 לתנועה ימינה או שמאלה ביחס לתנועה המקורית.
- תקבלו r העלות של ביצוע צעד.
- לסיכום:
- W, H, L, p, r