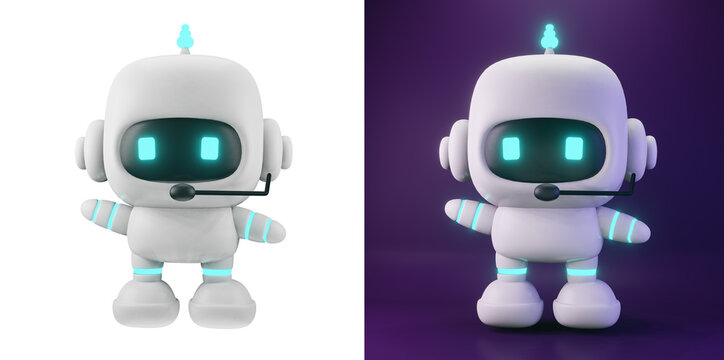
**תרגיל בית 2 מבוא לבינה מלאכותית**

**חיפוש רב סוכנים**

**בינה מחסנים**



**מגישים:**

**יגל מימון – 316611078**

**יורם פרטוש - 209700103**

**הקדמה ואדמיניסטרציה**

**הנחיות כלליות**

* תאריך הגשת התרגיל: **24.3.2024**
* את המטלה יש להגיש בזוגות בלבד – בקשות להגשה ביחידים באישור המתרגל האחראי בלבד (ספיר טובול).
* יש להגיש את המטלה מוקלדת בלבד – פתרון בכתב יד לא ייבדק.
* התשובות צריכות להיות כתובות בשפה העברית או באנגלית.
* אפשר לשלוח שאלות בנוגע לתרגיל דרך הפיאצה.
* המתרגלת האחראית על התרגיל: **אופק גוטליב**.
* בקשות דחיה מוצדקות יש לשלוח למתרגל האחראי בלבד.
* במהלך התרגיל ייתכן שיעלו עדכונים – הודעה תפורסם בהתאם במקרה זה.
* העדכונים מחייבים וזוהי אחריותכם להתעדכן לגביהם עד מועד הגשת התרגיל. עדכונים יופיעו בטופס בצבע צהוב.
* העתקות תטופלנה בחומרה.
* ציון המטלה כולל חלק יבש וחלק רטוב. בחלק היבש נבדוק שתשובתכם נכונה, מלאה, קריאה ומסודרת. בחלק הרטוב הקוד שלכם ייבדק הן על הגבלת הזמן שתפורט בהמשך ועל אחוזי ההצלחה של האלגוריתמים שלכם לעומת אלו שאנו נממש כבדיקה.
* מומלץ להסתכל בקוד בעצמכם. שאלות בסיסיות על פייתון שלא נוגעות לתרגיל כדאי לבדוק באינטרנט לפני שאתם שואלים בפיאצה. מומלץ לקרוא את הקוד הנתון על מנת להבין את אופן פעולתו – במקרה שישנם דברים לא מובנים . (לשם כך יש הערות רבות ואף הסבר מורחב על הסביבה!)
* מומלץ לא לדחות את התרגיל לרגע האחרון מאחר שהמימוש וכתיבת הדו"ח עלולים לקחת יותר זמן מהצפוי.
* התייחסות בלשון זכר, נקבה או רבים מתייחסים כלפי כלל המינים.
* אין צורך להשתמש בתהליכים ובמאפיינים מורכבים של מערכות הפעלה, התרגיל בית אמור לרוץ על כל מערכת הפעלה ולא למערכות הפעלה ספציפיות. אם ישנה בעייה פרטנית, שלחו מייל למתרגלת האחראית על התרגיל.

**הוראות הגשה**

בתוך קובץ זיפ עם השם : HW2\_AI\_id1\_id2.zip

את הדו"ח היבש בפורמט הבא : id1\_id2.pdf

ואת הקובץ: submission.py שבו אתם ממשים את האלגוריתמים

**הקדמה**

בינה מחסנים רוצים להפוך את המחסן לאוטונומי והם מתלבטים מיהו הרובוט אותו הם רוצים לשכור למשימה. על הרובוטים רובוט R1 ו-R2 להתחרות אחד בשני והמנצח יתקבל לעבודה.

בתרגיל זה תממשו ותחקרו אלגוריתמי משחק סכום אפס אדברסריאלים שלמדתם בהרצאות ובתרגולים.

**תיאור המשחק**

המשחק מתרחש בלוח משבצות בגודל 5X5 כאשר על הלוח: 2 רובוטים, 2 תחנות הטענה,   
2 חבילות ו-2 יעדים (אחד לכל חבילה).   
לכל רובוט יחידות הטענה (battery), וניקוד (credit).

מטרת כל רובוט לצבור יותר ניקוד מהרובוט האחר עד סוף המשחק, המשחק נגמר כאשר לאחד מהרובוטים נגמרת הסוללה או כאשר נגמר מספר הצעדים המקסימלי לכל רובוט (ערך מוגדר מראש, דוגמה בהמשך).

ניקוד נצבר כאשר רובוט מביאה חבילה אל היעד שלה. כאשר רובוט מביאה חבילה ליעד הוא מקבל N יחידות ניקוד כאשר N הוא מרחק מנהטן בין מיקום החבילה המקורי ויעדה של החבילה כפול 2. לאחר שחבילה מגיעה אל היעד שלה היא והיעד שלה נעלמים ובמקומם מופיעים יעד וחבילה חדשים. כלומר, בכל עת על הלוח יש בדיוק שתי חבילות ושני יעדים, אחד לכל חבילה.

רובוטים מתחילים עם טעינה וללא חבילה. כל פעולת תנועה של הרובוט עולה לו יחידת סוללה אחת. רובוט יכול לנוע למעלה, למטה, ימינה, שמאלה. בנוסף, רובוט יכול להטעין את הסוללה בתחנת הטענה כאשר הוא נמצא במשבצת של תחנת הטענה, הוא עושה זאת ע"י המרה של כל יחידות הניקוד שלו ליחידות טעינה. רובוט יכול לאסוף ולהוריד חבילה. רובוט אוסף חבילה כאשר הוא עומד באותה משבצת כמוהה ומבצע פעולת אסיפה. רובוט מוריד חבילה כאשר הוא נמצא ביעד ומבצע פעולת הורדה. (אי אפשר להוריד חבילה שנאספה במשבצת שאינה היעד)

סה"כ לרובוט 7 פעולות אפשריות בכל עת:

move north, move south, move east, move west, pick up, drop off, charge

לא יתאפשר צעד למשבצת לא חוקית (משבצת לא חוקית היא משבצת מחוץ לגבולות הלוח או כזו שנמצא בה הרובוט השני). לא יתאפשר הטענה, אסיפה, הורדה במשבצות לא חוקיות. כל סוכן יכול להטעין בכל אחת מהתחנות הטענה.

**הסבר על הסביבה ו הרצת / דיבוג המשחק**

בשונה מתרגיל בית 1 הפעם תעבדו עם קבצי py ולא מחברות pynb. מוזמנים לעבוד על התרגיל בכל IDE מתאים שנוח לכם לעבודה בפייתון, אנו ממליצים על pycharm של חברת jetBrains אליה יש לכם מנוי מטעם הטכניון.

הסביבה שאיתה תעבדו ממומשת בקובץ WarehouseEnv.py מוזמנים לעיין בה.

בקובץ Agent.py ממומשים סוכנים מהם הסוכנים שאתם תממשו יורשים, מומלץ להסתכל על הסוכנים הממומשים בה בכדי להבין כיצד הם עובדים עם הסביבה.

מומלץ להסתכל על הפונקציות get\_legal\_operator ו - apply\_operator בכדי להבין את אופן ההתממשקות שלכם עם הסביבה.

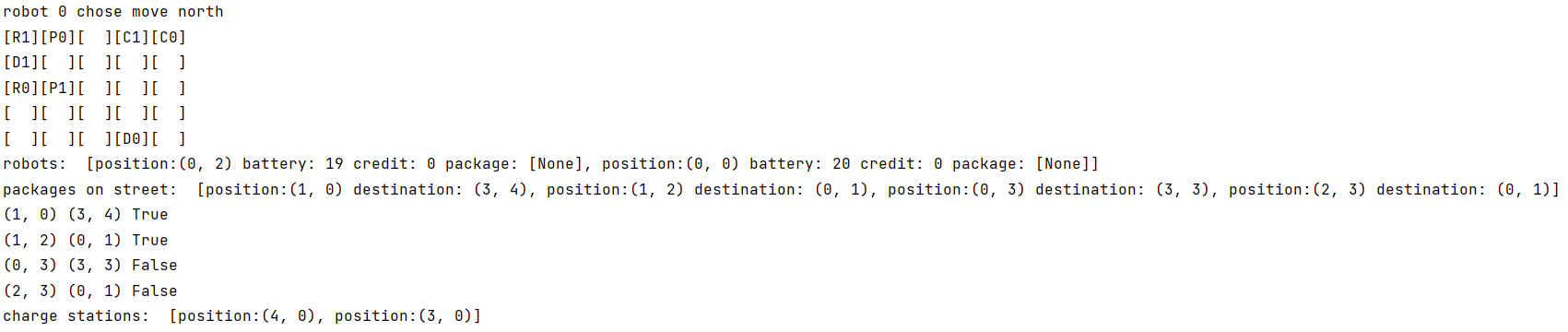
שימו לב שבקובץ submission יש סוכן שנקרא hardcoded, מטרתו לעזור לכם להבין את הסביבה. לפני שאתם שואלים לגבי כיצד הסוכן יתנהג אם יבצע פעולה בדקו בעצמכם בעזרת סוכן זה.

**הרצה**

הרצת משחק נעשית ע"י שורת הפקודה הבאה שמריצים מהטרמינל:

python main.py greedy random -t 0.5 -s 1234 -c 200 --console\_print --screen\_print

* הארגומנט הראשון (במקרה הזה greedy) מציין את האלגוריתם שלפיו ישחק agent0
* הארגומנט השני (במקרה הזה random) מציין את האלגוריתם שלפיו ישחק agent1
* הגבלת זמן לצעד t- מקבלת ערך שמייצג את מספר השניות המקסימילי לצעד
* גרעין לקבלת ערך רנדומלי s- מקבל ערך שעוזר לחולל סביבה באופן רנדומלי, כאשר מועבר אותו ערך seed תחולל אותה סביבה
* מספר הצעדים המקסימלי עבור סוכן c-
* ערך console\_print-- דגל אופציונלי, כאשר מועבר מתבצעת הדפסה לקונסולה של המשחק שנראת כך :

מודפס מספר הסוכן יכול להיות 0 או 1 והאופרטור בו בחר. לאחר מכן מודפס הלוח לאחר שהסוכן הפעיל את האופרטור. הלוח כולל 25 משבצות כאשר בכל אחת יכולים להיות רובוט, חבילה המחכה לרובוט או תחנת הטענה. 

האותיות אשר מופיעות בלוח מסמלות:

R – Robot

C – Charge station

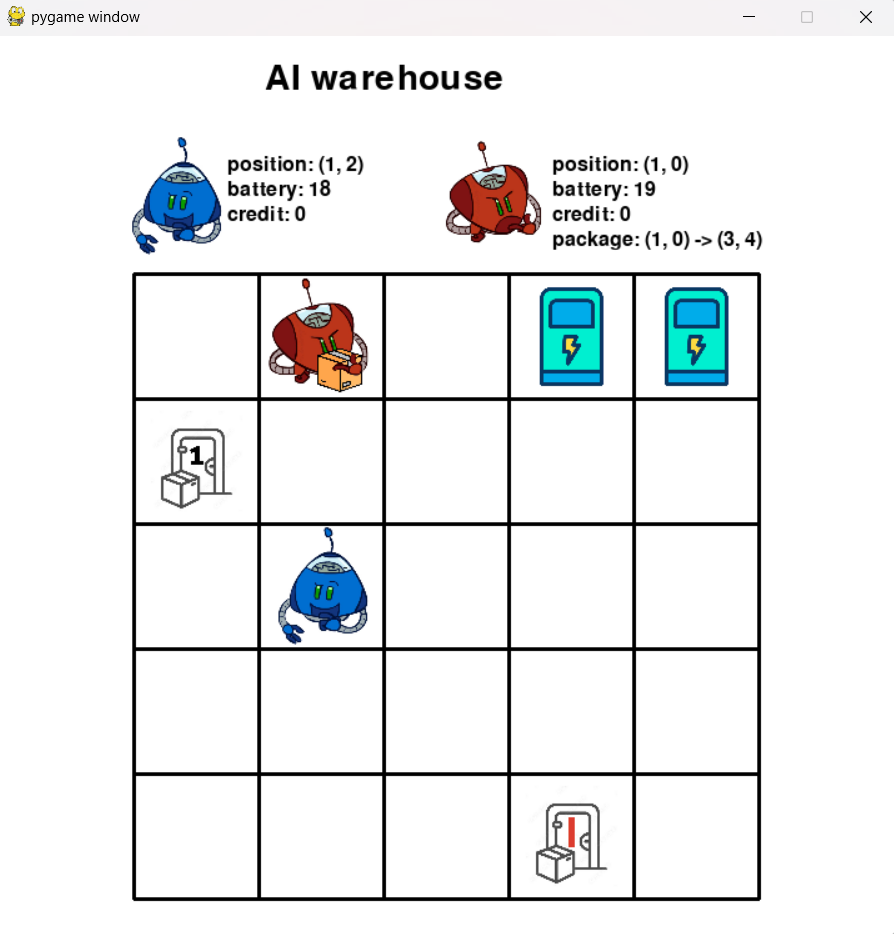
P – Package on street

D – Destination of Package on street

X – Destination of Package taken by Robot

המספר המופיע ליד כל אחת מהאותיות הוא המזהה של האובייקט אליו הוא משתייך – לדוגמא: R1 עבור הרובוט הראשון. עבור חבילה שנאספה ע"י רובוט, המספר ליד X הוא המזהה של הרובוט שאסף אותה. עבור חבילה במרחב, המספר ליד D הוא המזהה של החבילה במרחב. לעיתים מספר אובייקטים מופיעים באותו המיקום בתמונה ואז מודפס רק אחד מהם. במקרה זה, אפשר להשתמש ברשימות המפורשות שמופיעות אחרי הלוח. בנוסף שימו לב! מופיע לכם ברשימות את המקור והיעד של שתי החבילות הבאות שיופיעו על הלוח, מוזמנים ומומלץ להשתמש במידע זה בהמשך התרגיל.

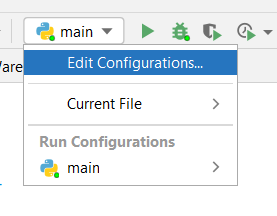
* ערך screen\_print-- דגל אופציונלי, כאשר מועבר יודפס לכם הנפשה של המשחק בחלון pygame שנראה כך:

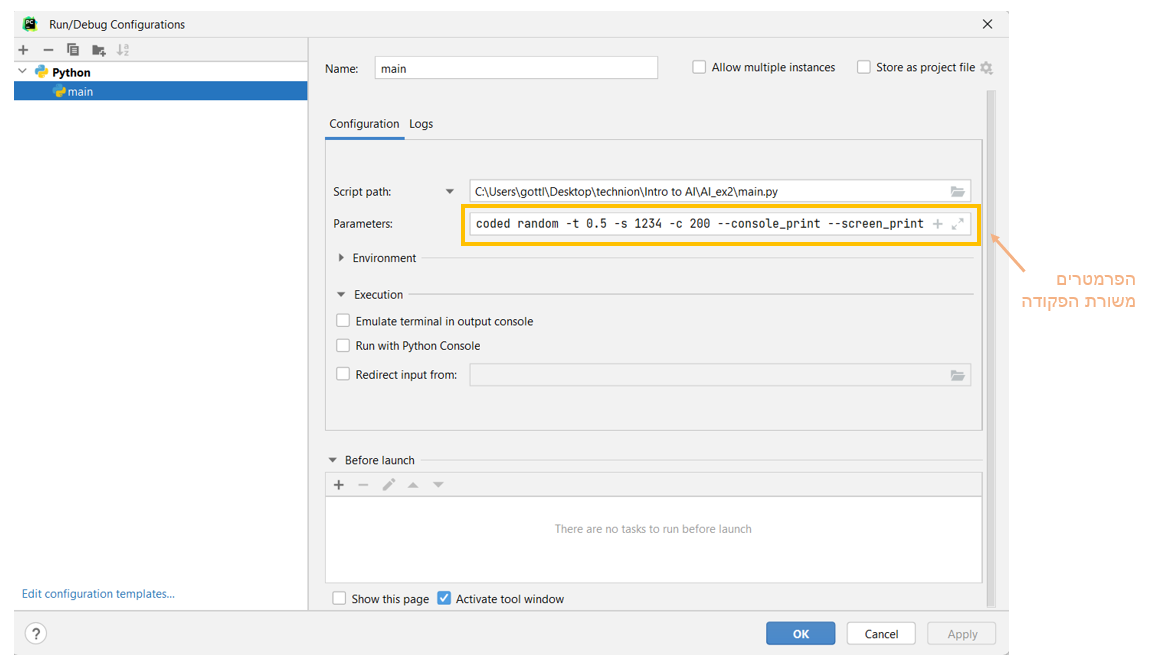


בנוסף בכדי לבדוק את הסוכנים שלכם על מספר רב של משחקים, אתם יכולים להוסיף את הדגל tournament-- וכך תקבלו הרצה רצופה של n משחקים. מוזמנים לשחק עם המשתנה של מספר המשחקים ולראות האם מקבלים התנהגויות רצויות עבור האלגוריתמים.

**דיבוג**

בכדי לדבג את המשחק עליכם לעקוב אחר השלבים הבאים:

2.



**מתחילים לכתוב !**

**חלק א - ImprovedGreedy**

1. (יבש: 4 נק') כפי שלמדתם, אנו מגדירים בעייה במרחב בתור רבעייה (S, O, I, G). הגדירו פורמלית (הסבירו כיצד נראים ערכי הרבעייה) את המשחק המתואר לכם ע"פ הנתונים שאתם מקבלים מהסביבה.

S – מרחב המצבים. במקרה שלנו מרחב המצבים הוא רשימה של 7 משתנים שמיוצג כך:

S = (R0\_loc, R1\_loc, battery0, battery1, credit0, credit1, turn, steps\_remain, charger0\_loc, charger1\_loc, package0\_loc, package1\_loc, goal0, goal1)

כאשר:

RX\_loc – מספר בין 0 ל25 שמייצג את מיקומו של רובוט X כאשר X יכול להיות 0 או 1.

Turn- הרובוט שתורו עכשיו (0 או 1).

batteryx – מספר אי שלילי המייצג רמת הטענה עבור רובוט x כאשר x יכול להיות 0 או 1.

creditx – מספר אי שלילי המייצג הניקוד שצבר הרובוט x כאשר x יכול להיות 0 או 1.

steps\_remain- מספר הצעדים שנותרו שזה שווה למספר שהכניס המשתמש פחות מספר הצעדים שנעשו כבר.

chargerX\_loc- מספר בין 0 ל25 שמייצג את מיקומו של עמדת ההטענה X כאשר X יכול להיות 0 או 1.

packageX\_loc - מספר בין -1 ל25 שמייצג את מיקומו של החבילה ה X כאשר X יכול להיות 0 או 1(ועבור -1 אם החבילה נאספה).

goalX- מספר בין 0 ל25 שמייצג את מיקומו של מיקום היעד עבור חבילה X כאשר X יכול להיות 0 או 1.

O – מוגדר להיות מרחב המצבים כאשר נשים לב שמרחב המצבים מתחלק לפעולה עבור רובוט 0 ורובוט 1, בכל פעולה מהפעולות המוזכרות במסמך זה הרובוט השני מוכרח לחכות לכן מרחב הפעולות יהיה:

O = {O0, O1}

כאשר:

OX = {𝑚𝑜𝑣𝑒 𝑛𝑜𝑟𝑡ℎ,𝑚𝑜𝑣𝑒 𝑠𝑜𝑢𝑡ℎ,𝑚𝑜𝑣𝑒 𝑒𝑎𝑠𝑡,𝑚𝑜𝑣𝑒 𝑤𝑒𝑠𝑡,𝑝𝑖𝑐𝑘 𝑢𝑝,𝑑𝑟𝑜𝑝 𝑜𝑓𝑓,𝑐ℎ𝑎𝑟𝑔𝑒, Wait}

כאשר X יכול להיות 0 או 1 ומוגדר להיות מרחב הפעולות עבור רובוט X.

I – מוגדר להיות מרחב המצבים ההתחלתי. על פי הסקלטון שקיבלנו מהסגל והגדרת המשחק במסמך מתקיים שלכל רובוט רמת סוללה 20 וקרדיט 0 בתחילת המשחק.

לכן:

I = {

RX\_loc – מספר רנדומלי בין 0 ל25.

Turn- 0 או 1 באופן רנדומלי.

batteryx – המספר הקבוע 20.

creditx – המספר הקבוע 0.

steps\_remain- מספר הצעדים שקיבלנו מהמשתמש.

chargerX\_loc מספר רנדומלי בין 0 ל25.

packageX\_loc - מספר רנדומלי בין 0 ל25.

goalX- מספר רנדומלי בין 0 ל251.

}

G – קבוצת מצבים סופיים שעל פי המסמך צריך להיות או כאשר מספר הצעדים שהוגדר על ידי המשתמש נגמר או כאשר **אחד מהרובוטים** מסיים את הסוללה לכן:

1. (יבש: 4 נק') הגדירו היוריסטיקה משלכם להערכת מצבי המשחק. עליכם לתעד אותה בנוסחה מפורשת ועלייה לכלול לפחות שלושה מאפיינים של הסביבה. בחרו שמות ברורים בנוסחה שלכם:

ראשית נסתכל על הנתונים במצב כלשהו שמוגדרים כמועילים ביותר עבור הרובוט.

נשים לב שמיקום כלשהו לא משמעותי ללא יחס למיקום של אובייקט אחר, לכן:

RX\_loc יקבל משמעות רק כאשר נשכלל אותו עם PackageX\_loc או עם goalX (תלוי בערך packageX\_coll) (אנחנו מעריכים שיש משמעות זניחה עבור מיקום הטעינה מאחר והמשחק נגמר ברגע שאחד הרובוטים מסיים את הסוללה לכן.

נתייחס ליחסים אלו כמרחק מנהטן מאחר וזה מרחק שיותר מייצג את המציאות.

בנוסף כל רובוט יירצה לסיים את הסוללה כאשר הוא מוביל.

וכן אין משמעות chargerX\_loc אם batteryx גדול מ steps\_remain.

לכן נבדוק האם (steps\_remain > batteryx and creditx < credity) כאשר y זה הרובוט השני.

לכן היוריסטיקה שלנו תראה כך:

1. (רטוב: 10 נק') ממשו בקובץ submission.py את הפונקציה **smart\_heuristic**

שבה משתמש הסוכן AgentGreedyImproved

בקוד

1. (יבש: 2 נק') מהו החיסרון העיקרי של האלגוריתם? (לעומת minimax)

האלגוריתם החמדן פחות מיודע ומתייחס רק לשלב הבא.

למשל אם בעוד 2 שלבים המצב לא לטובת השחקן אבל בשלב הבא המצב כן לטובתו האלגוריתם יבצע צעד שיעדיף את השלב הבא האופטימלי מאשר זה שבעוד 2 שלבים.

**חלק ב - RB-Minimax**

1. (יבש 3 נק') מה היתרונות והחסרונות של שימוש בהיוריסטיקה קלה לחישוב לעומת היוריסטיקה קשה לחישוב בהינתן שהיוריסטיקה הקשה לחישוב יותר מיודעת מהקלה לחישוב ? בהינתן שאנו בmin-max מוגבל משאבים.

ביוריסטיקה קלה לחישוב אני יכול לרדת יותר בעץ וכך להגיע יותר עמוק אבל היא פחות מיודעת לעומת זאת ביוריסטיקה קשה לחישוב יותר מיודעת

1. (יבש: 4 נק') חברתכם לקורס דנה מימשה סוכן minimax, היא שמה לב כי לעיתים הסוכן יכול לנצח בצעד אחד אך הוא בוחר בצעד אחר. האם יש לה באג באלגוריתם? אם אין באג הסבירו מה באלגוריתם גורם להתנהגות שכזו. אם יש באג מה הוא יכול להיות?

לדנה אין באג כי ייתכן שהיוריסטיקה ממזערת הפסד עד עומק מסויים ולא הסתכלה עד סוף המשחק ולכן השובר שוויון שם בחר בדרך לא נכונה למרות שאם היה בוחר בדרך אחרת אולי היה מנצח.

1. (רטוב: 10 נק') עליכם לממש את המחלקה AgentMinimax בקובץ submission.py. שימו לב! הסוכן מוגבל משאבים, כאשר המשתנה time\_limit מגביל את מספר השניות שהסוכן יכול לרוץ לפני שיחזיר תשובה. (הגבלת הזמן עלייה אתם נבדקים הינה שנייה כלומר 1 t-).

בקוד.

1. (יבש: 3 נק') נניח שבסביבה היו K שחקנים במקום 2. אילו שינויים יהיה צריך לעשות במימוש סוכן Minimax? **כתבו פסאודו קוד בדומה לזה שראינו בתרגול.**
   1. בהינתן שכל סוכן רוצה לנצח ולא אכפת לו רק ממכם.

מאחר וכל אחד רוצה לנצח בלי קשר לאחרים (קצת כמו גרידי) אז הוא לא מתחשב בשאר הסוכנים.

* 1. בהנחה והדבר היחיד שכל סוכן רוצה הוא שלא תנצחו.

נעריך ששאר השחקנים ירצו לשחק כשחקן מינימום להביא את היוריסטיקה שלנו למינימום ועם זאת נרצה בתורנו להגיע למצב מקסימאלי.

* 1. בהנחה שכל סוכן רוצה שהסוכן שאחריו בתור ינצח.

בדומה לסעיף א׳ רק שהפעם במקום למקסם את המצב שלו עצמו הסוכן ימקסם את המצב של הסוכן שאחריו.

**חלק ג - Alpha-Beta**

1. (רטוב: 10 נק') ממשו שחקן אלפא - בטא מוגבל משאבים במחלקה AgentAlphaBeta בקובץ submission.py, כך שיתבצע גיזום כפי שנלמד בהרצאות ובתרגולים.

בקוד

1. (יבש: 3 נק') האם הסוכן שמימשתם בחלק זה יתנהג שונה מהסוכן שמימשתם בחלק ב מבחינת זמן ריצה ובחירת מהלכים ?

הסוכן אלפא-ביתא יתנהג שונה מבחינת זמן ריצה ומהלכים בגלל ששני הסוכנים מתנהגים דומה ויש להם הגבלה בזמן הריצה עבור (עד שנייה חישוב לכל צעד) אבל ההבדל החשוב הוא שאלפא בתא הוא מתעלם מפיתוחים מיותרים לכן הגבלה זו לכן הוא מצליח לפתח יותר בזמן המוגבלה שניתן לו ולהגיע עמוק יותר בעץ, לכן זמן הריצה שלו קטן יותר כי חוסך פיתוחים מיותרים ומאחר והוא לוקח בחשבון יותר נתונים הוא גם מתנהג אחרת (יכול לבחור צעד שמוביל למצב עמוק בעץ שמינימקס לא מודע אליו).

**חלק ד - Expectimax**

1. (יבש: 3 נק') בהנחה ואתם משתמשים באלגרותים Expectimax נגד סוכן שמשחק באופן רנדומלי לחלוטין באיזה הסתברות תשתמשו?

נשתמש בהסתברות יוניפורמית כי הסוכן הרנדומלי בוחר את הצעד הבא באופן רנדומלי לחלוטין מרשימת הצעדים האפשריים

1. (יבש: 4 נק') עבור משחקים הסתברותיים כמו שש בש, בהם יש מגבלת משאבים, משתמשים באלגוריתם .RB-Expectimax הניחו כי ידוע שהפונקציה היוריסטית ℎ באלגוריתם Expectimax-RB מקיימת

איך ניתן לבצע גיזום לאלגוריתם זה? תארו בצורה מפורטת את תנאי הגזימה, והסבירו את הרעיון מאחוריו.

אפשר לגזום את שאר הבנים בכך שאם נקבל מתישהו מאחד הבנים את הערך 1 אם מדובר בסוכן מקסימום נגזום את שאר הבנים מאחר ולא ייתכן שנמצא ערך גדול מ1.

בנוסף עבור סוכן מינימום ניתן לגזום את שאר הבנים אם נקבל את הערך -1 מאותה סיבה (ההופכי לה)

1. (רטוב: 10 נק') הסוכן של minimax ו-alpha-beta מניח שהסוכן היריב יבחר באופרטור שיוביל לתוצאה האופטימלית בתורו, אולם זה לא תמיד מתרחש.

לדוגמה, כאשר אנו מתחרים עם סוכן חמדן, סביר להניח שהוא לא יבחר בפעולה האופטימלית בכל צעד. אפשר להתחשב באפשרות שהיריב יבחר בפעולה שאינה אופטימלית בתורו באמצעות סוכן Expectimax.

גילינו מידע סודי על הרובוט המתחרה, הוא בוחר בין כל הפעולות בצורה יוניפורמית (בצורה אחידה) אבל לתזוזה ימינה ולאסיפת חבילה (כאשר פעולות אלו אפשריות) יש הסתברות גדולה פי 2 מלשאר הפעולות.

ממשו אלגוריתם Expectimax המשתמש במידע הסודי שקיבלתם.

**חלק ה - משחק עם פקטור סיעוף גדול**

1. (יבש: 6 נק') להלן שינויים אפשריים ששוקלים בינה מחסנים לעשות במשחק בכדי לבחון יכולות נוספות של הרובוטים. עבור כל שינוי ציינו מה ההשפעה שלו על מקדם הסיעוף וחשבו את מקדם הסיעוף החדש המתקבל.
   1. הגדלת לוח המשחק להיות 8X8 והוספת מחסומים בסביבה. (מחסומים משמע משבצות שהסוכן לא יכול לעבור בהן)
   2. הוספת היכולת של רובוט בכל תור לבחור משבצת על הלוח ולהניח עלייה בלוק, משמע בכל תור יכול הרובוט לנוע למעלה, למטה,ימינה, שמאלה, לאסוף חבילה, להוריד חבילה, להטעין, ולהניח בלוק על הלוח, בלוק יכול להיות מונח על כל משבצת ריקה.
2. (יבש: 6 נק') בהנחה ומימשו את השינוי השני עבור הסביבה (סעיף 1b)
   1. האם יש אלגוריתם מהסעיפים הקודמים שנוכל להשתמש בו שזמן הריצה שלו סביר? (סביר משמע לא גדול מהותית מהזמן שלוקח לו להחזיר צעד עבור המשחק בלי השינוי).
   2. הציעו אלגוריתם שונה מאלו שממשתם בסעיפים הקודמים שנלמד בקורס שירוץ בזמן סביר. הסבירו מדוע בחרתם בו ולמה הוא טוב להתמודדות עם האתגר שנוצר משינוי הסביבה.

**חלק ו - יבש - שאלה פתוחה - MCTS**

שחקן אדום ושחקן כחול שיחקו משחק. להלן עץ המשחק שמתאר את העץ שנוצר בשלב ביניים בהרצת MCTS עם פיתוח צמתים לפי UCB1 על משחק סכום אפס בין שניהם, נתון .

a

b

c

43/100

1/9

6/21

3/4

5/5

0/3

10/20

6/7

דגש על האלגוריתם: הערך בצומת הכחול מייצג את כמות הניצחונות של השחקן האדום מתוך כמות המשחקים שבוצעו עם הצעד הזה (ולהיפך). למשל צומת b מייצג את כמות הניצחונות של השחקן האדום אם בחר בפעולה שגרמה לו להגיע למצב b מתוך סך כל המשחקים ששוחקו עם הצעד.

1. (5 נק') חלק מהערכים בצמתים נמחקו, השלימו את החלקים החסרים, אין צורך לנמק.
2. (5 נק') הצומת הבא שייבחר בשלב ה - selection יהיה (הוסיפו חישובים לנמק את בחירתכם):
   1. צאצא של a
   2. צאצא של b
   3. צאצא של c
3. (5 נק')בהנחה שכל סימולציה מכאן והלאה מסתיימת בניצחון של השחקן הכחול מה מספר הניצחונות המינימלי שנדרש כדי שצאצא אחר של השורש ייבחר בשלב ה-selection (הוסיפו חישובים לנמק את תשובתכם)?
4. (3 נק') כעת רוצים לבצע שינוי כך שנעדיף exploration יותר מ - exploitation. הגישה לנוסחה שמחשבת את ה - UCB1 חסומה לכם, אך הנוסחה משתמשת ב - אשר אליו יש לכם גישה ואתם יכולים לשנותו. כיצד תשנו אותו בכדי שהנוסחה החדשה שתיווצר תעדיף יותר exploration מ - exploitation לעומת הנוסחה הקודמת.