תרגיל סיום – נושאים בבינה מלאכותית ולמידה מבוססת חיזוקים

תאריך הגשה 23.2

לצורך תרגיל זה כדאי

* 1. לדאוג שהחישובים שלכם יעילים ומהירים כדי שתוכלי להריץ הרבה משחקים.
  2. להיות סבלניים, ולתת זמן לרשת להתאמן. לכן, כדאי שתדאגו שכל תהליך האימון שמתואר למטה יהיה אוטומטי, ותתנו לו לרוץ בלילות על המחשב שלכם. זה לוקח זמן.
  3. ללמוד עצמאית על רשתות נוירונים. לצורך כך אני ממליץ לצפות בסרטון הבא:

<https://www.youtube.com/watch?v=aircAruvnKk&t=5s>

וכן בזה שאחריו. הם מסבירים יפה את עיקרי המודל. הסרטונים הבאים מסבירים את האלגוריתם לאימון רשת נוירונים. ידע חשוב, אבל לא הכרחי לטובת התרגיל, שכן אתם יכולים להשתמש במימוש קיים. לאחר מכן, תראו סרטון על שימוש בספריית pytorch לכתיבת רשת. יש סרטונים של כ15 דקות שמראים איך לממש. מי שרוצה להבין לעומק יש סדרה של סרטונים על איך לממש רשת נוירונים ולאמן אותה בפייטון. זה ידע חשוב באופן כללי, אבל הספריה מייתרת את רוב העבודה ומאפשרת לכם לעשות את הכל במהירות. היא גם משתמשת בקוד הרבה יותר יעיל ממה שתוכלו לכתוב בעצמכם ולכן האימון יהיה הרבה יותר מהיר (אבל עדיין ייקח הרבה זמן). אז כאמור, התכנות בפייטון בשימוש בספרית pytorch.

1. ממשו רשת נוירונים שמנסה ללמוד את הפונקציה היוריסטית שלכם. הפונקציה תקבל דוגמאות מהצורה (לוח, ערך) ותנסה ללמוד לנבא את הערך היוריסטי. יצרו דוגמאות שונות של לוחות (למשל על ידי מספר משחקים רנדומיים) והשתמשו בפונקציה היוריסטית שלכם כדי לחשב את הערך של כל לוח. פונקציית המטרה תיהיה mean squared error. כלומר, עליכם למצוא משקולות לרשת כך מינימלי. כאן הוא הערך שהרשת נותנת ללוח x, ואילו h(x) הוא הערך שהיוריסטיקה נותנת ל-x. כלומר, הרשת מנסה שההפרש בין מה שהיא אומרת לבין מה שהיוריסטיקה אומרת יהיה קטן.
   1. מטרת החלק הזה היא ללמוד לעבוד עם רשת נוירונים ולאתחל רשת לחלק השני. השתמשו בספריית pytorch. אתם יכולים להעזר בChatGPT ודומיו על מנת להבין איך לממש את זה וגם אחד בשני. שימו לב שבבדיקה הפרונטלית אבדוק הבנה של הקוד. אז אין בעיה למצוא קוד ברשת, אבל צריך להבין מה עשיתם.
   2. דאגו לנרמל את הערך של היוריסטיקה להיות בין 0 ל-1, כאשר 1 כמובן מעיד על סיכוי גבוה לניצחון ו-0 על סיכוי גבוה להפסד. כדאי גם לנסות לנרמל את תאור הלוח כך שכל כניסה בוקטור תהיה בין 0 ל 1.
   3. שכבת הקלט של הרשת תהיה ייצוג של הלוח. אתם יכולים לבחור ייצוג כרצונכם. אבל, בדרך כלל רצוי שהערכים של כל רכיב בוקטור הקלט יהיו בין 0 ל-1. לשם כך, אתם יכולים לשחק עם הייצוג כרצונכם ולהוסיף עוד כניסות.
   4. לרשת תהיינה לפחות שתי שכבות fully-connected בגודל 40 לפחות כל אחת.
   5. לרשת שכבת פלט שמכילה נוירון בודד שמחזיר ערך בין 0 ל-1. מומלץ בשכבה הזאת לא להשתמש בפונקציה לא-לינארית (non-linear activation function).
2. אמנו רשת ישירות לנבא את הערך של מצב בצורה הבאה:
   1. התחילו עם הרשת שאימנתם בשלב 1. (לא חובה, אפשר גם רשת עם משקלות התחלתיים רנדומליים).
   2. הריצו מספר משחקים לדוגמה תוך שאתם משתמשים בשחקנים יוריסטיים. את המהלך של השחקן תבחרו בצורה הבאה: מצאו את 4 המהלכים הטובים ביותר לפי היוריסטיקה. בחרו בניהם אקראית כשתתנו לכל מהלך משקל ev כאשר v הערך היוריסטי של המהלך (כלומר הלוח שתקבלו בסופו). כמובן, צריך לנרמל את המשקלות על ידי חלוקה בסכום המשקלות של ארבעת המהלכים.
   3. שימרו את הלוחות שקיבלתם לאורך המשחק והשתמשו בהם כדוגמאות לצורך למידה כפי שעשיתם ב1. אלא שכעת, אם הלוח מייצג מצב שהתקבל על ידי מהלך של השחקן המנצח, תנו לו ערך 1. אם הלוח התקבל על ידי מהלך של השחקן המפסיד, תנו לו ערך 0. אפשר גם לנסות לשפר – למשל, הציון יהיה קשור לאופי הנצחון. אם עשיתם מרס, אז 1, אם ניצחתם בהפרש גדול (כלומר, ליריב היו עוד הרבה שחקנים) אז נניח 0.9 ואם הנצחון היה ממש ברגע האחרון, אז 0.6. כמובן, המספרים הללו הם רק לדוגמה. הרעיון הוא לתת לרשת מידע יותר מדויק על ההצלחה שלכם, כדי שתנבא בעתיד יותר טוב.
   4. שפרו את הרשת הנוכחית בעזרת הדוגמאות שייצרתם.
   5. חזרו על שלב b מספר פעמים. בכל פעם כשאתם חוזרים עליו, השתמשו ברשת הנוכחית כדי לחשב את הערך היוריסטי של המצב. בידקו בדרך שאכן יש שיפור – למשל, תנו לרשת החדשה לשחק נגד הרשת הישנה, ותראו שהיא אכן מנצחת אותה. (כמובן, חשוב להריץ מספר משחקים).
      1. שימו לב – כדי להגיע לשחקן ממש טוב, כנראה תצטרכו לאמן את הרשת הרבה זמן, ולכן, יהיה מספיק אם תריצו מספר לא גדול של שלבים אבל תראו שיש התקדמות בין שלב לשלב. אמנם יהיה בונוס למי שהשחקן שלו יהיה חזק, אבל זה לא יהיה העיקר.
3. שיפורים אפשריים: בכל שלב, השתמשו בMCTS כדי לבחור את הפעולה של השחקן כשאתם משתמשים ברשת כדי לתת ערך יוריסטי לעלים בעץ.
4. הגשה וציון: יש להגיש דו״ח ביחד עם סרטון באורך של עד 10 דקות המסביר את מה שעשיתם, את מבנה הרשת שלכם, ומבנה הקוד שלכם, וגם מציג תוצאות שונות.
   1. חלק א׳: 20%
      1. מעבר לתעוד, הריצו 4 או יותר משחקים מול שחקן רנדומי. לכל אחד מ50 הלוחות (או פחות, אם נצחתם קודם) חשבו את הערך היוריסטי ואת הערך שהרשת נותנת (עליכם להגיע ל200 לוחות סה״כ, אם צריך יותר משחקים, אז הריצו עוד משחקים) ואת הממוצע והשונות של ההפרש בין הערך של היוריסטיקה והערך של הרשת.
   2. חלק ב׳: 80%
      1. בסיום כל שלב אימון (כלומר, עדכון הרשת הנוכחית עם הדוגמאות) הריצו 10 משחקים מול שחקן שמשתמש ביוריסטיקה המקורית שלכם כדי לבחור פעולה וחשבו את הניקוד הממוצע של השחקן מבוסס הרשת. ציירו גרף המתאר ערך זה כפונקציה של מספר שלבי האימון.
   3. שימו לב – תתקיים גם בדיקה פרונטלית, וכן טורניר בין השחקנים השונים, כפי שצוין בשיעור הראשון. כפי שכתוב למעלה, נתחשב במגבלות זמן האימון גם.