עבודת תכנות 3

1. פתרו את בעיית mountain-car בעזרת שימוש באלגוריתם policy gradient ושימוש בfunction approximation-.
   * כfunction-approximator- השתמשו בשיטת tile-coding שתוארה בשיעור.
   * אתם יכולים לבחור את שיטת הריצוף כרצונכם תוך שימוש ב-100 פיצרים לכל היותר.
   * אתם יכולים לבחור איזה אלגוריתם policy gradient להשתמש.
     1. אם בחרתם באלגוריתם מסוג actor-critic אז גם להערכת פונקצית הערך/Q/A עליכם להשתמש ב function approximator. גם כאן, עד 100 פיצרים בשיטת tile-coding. הפיצרים לא חייבים להיות זהים לאילו שהשתמשתם לצורך המדיניות
   * גאמה יהיה 0.95.
   * כל שאר הפרמטרים תבחרו כרצונכם.
   * הגשה:
     1. הגישו את הקוד שלכם מתועד
     2. צרפו הסבר בקובץ pdf על הבחירות של האלגורים
     3. צרפו גרף התקדמות של הלמידה.
        + ציר ה-x של הגרף הוא מספר צעדי האימון
        + ציר ה-y של הגרף הוא איכות המדיניות
        + בציר ה-x ההתקדמות תיהיה כל 25000 צעדים
        + עבור ערך x כלשהו, כדי לחשב את ערך y המתאים לו, אתם תריצו את המדיניות שקיבלתם לאחר x צעדים 1000 צעדים, ותציינו את מספר האפיזודות שהרצתם. בהתחלה, סביר להניח שלא תצליחו, ולכן הערך יהיה סביב 5 כיוון שאורך האפיזודות חסום על ידי 200.
        + החישוב של ערך y עבור הגרף איננו חלק מתהליך הלמידה, ובצעדים הנ״ל לא יבוצע שינוי כלשהו במדיניות או בפרמטרים אחרים, אלא רק ישמש כדי להעריך את המדיניות הנוכחית.
   * הציון יתבסס על נכונות האלגוריתם והביצועים!