

ביולוגיה חישובית

תרגיל 3

זיהוי תבניות באמצעות רשת נוירונים

מגישים:

חיים הגר 318424900

עומר ארד 314096389

הוראות הרצה (יש להחליף את _ במספר 0 או 1):

- python3 buildnet_.py
- python3 runnet_.py

התוכנית מניחה כי הקבצים went_.py ו testnet_.py וקובץ nn_.txt תואם - קיימים בתיקיה ממנה מריצים את התוכנית.

פלט התוכנית runnet_.py הוא קובץ result_.txt

תיאור האלגוריתם: תחילה האלגוריתם בונה את הרשת ע"י קריאת קובץ הנתונים (סט האימון) והמרתו למילון, באופן כזה שהמחרוזת היא המפתח, והתיוג הוא הערך. לאחר מכן האלגוריתם הגנטי (שמקבל את המילון), מאתחל אוכלוסייה התחלתית של רשתות שונות כך שלכל רשת מוגרלים באקראי משקלים שונים (בטווח [1,-1]). לכל רשת מחשבים ציון כושר (fitness) ע"י חישוב היחס בין מספר התיוגים הנכונים (ע"י feedforward) לבין מספר המחרוזות הכולל, ובכל סיבוב מתבצעת בחירה של הרשתות בעלות הכושר הטוב ביותר על סט המבחן. את הרשתות הללו אנו מחלקים לזוגות, ומכל זוג רשתות כאלה האלגוריתם יוצר מספר רשתות חדשות במספר דרכים:

- דרך ראשונה היא על ידי חישוב ממוצע משקלים בין 2 הרשתות.
- דרך נוספת היא על ידי בחירת נקודת pivot, כך שהערכים עד נקודה זו נלקחים מרשת אחת, ומנקודה זו והלאה נלקחים הערכים מהרשת השנייה.

לאחר מכן, בהסתברות מסוימת מתבצע שינוי נוסף (מוטציה) ברשת החדשה שיצרנו. נבחרים בין 1 ל-3 קשתות באקראי ומגרילים מחדש את ערכן (בטווח [1,-1]).

ייצוג הדאטא: הנתונים מיוצגים בשלב האימון (בניית הרשת) על ידי מילון הממפה מחרוזות קלט לתגית. בשלב המבחן (הרצת סט מבחן על הרשת) הנתונים מיוצגים על ידי רשימת מחרוזות קלט.

מבנה הרשת: הרשתות (בשני הקבצים) מכילות 3 שכבות. שכבה ראשונה של 16 נוירונים, שכבה שנייה של 2 נוירונים ושכבה נוספת של נוירון בודד. כל נוירון בשכבה 1 מקושר לכל נוירון בשכבה 2. כל נוירון בשכבה 2 מקושר לנוירון בשכבה 3.

מאפייני האלגוריתם הגנטי: באלגוריתם הגנטי בחרנו ברשת הראשונה הסתברות למוטציה של 0.01, וברשת השנייה 0.02. בנוסף הגדרנו שלאחר 20 דורות ללא שיפור בציון הרשת נכריז על התכנסות. ציון הרשת (fitness) מחושב על ידי חישוב אחוז ההצלחה של האלגוריתם על סט האימון (חיזוי לכל קלט, השוואה לפלט האמיתי, ספירת הפעמים שצדקנו).

ביצועי התוכנית על קבוצת הלמידה ועל קבוצת המבחן: עבור 2 קבצי הקלט השונים הגענו לרמות דיוק של מעל 98% בשלבי האימון ובשלבי המבחן שערכנו (ראו נספח).

החוקיות:

- nn0.txt – בין 8 ל-12 אחדות (כולל) – תיוג 1. אחרת (7 ומטה או 13 ומעלה) – תיוג 0
- nn1.txt – עד 7 אחדות (כולל) – תיוג 1. אחרת (8 אחדות ומעלה) – תיוג 0

נספח:

פלט הרצה buildnet0:

```
Running, please wait...
0.9893020124186429 [#####] 100.00%
Best fit score: 0.9893020124186429
Generation: 100
Test score: 0.9890314569536424
Saving network...
DONE!
```

פלט הרצה buildnet1:

```
0.9903573030348333 [#####] 100.00%
Best fit score: 0.9903573030348333
Generation: 98
Test score: 0.9889583333333334
Saving network...
DONE!
```