**עבודה בביולוגיה חישובית – מטלה 3:**

דינה מיזלס ,319092201

הקוד נמצא בגיט - https://github.com/dinamaizlis/NeuralNetwork.git

במטלה זאת ממשתי רשת נוירונים עם אלגוריתם גנטי אשר יזהה תבניות על מחרוזות בינאריות באוך 16 ספרות.

**תכולת המטלה:**

* תבנית 0 :
  + **buidnt0** – מקבל traindata, testdat בונה רשת ומחזיר משקלים קובץ wnet0 (תוך שימוש באלגוריתם גנטי)
  + **runnet0** – מקבל data וקובץ משקלים wnet0 מריץ על הרשת ומחזיר קובץ testnet0 עם הסיווג label 0/1
* תבנית 1 :
  + **buidnt1** – מקבל traindata, testdat בונה רשת ומחזיר משקלים קובץ wnet1 (תוך שימוש באלגוריתם גנטי)
  + **runnet1** – מקבל data וקובץ משקלים wnet1 מריץ על הרשת ומחזיר קובץ testnet0 עם הסיווג label 0/1

**ייצוג הפתרון:**

**חלוקה לtrain/test:**

תחילה יצרתי פונקציה שתחלק את הדאטה nn0.txt & nn1.txt לשנים -train & test.

זאת כדי לפצל את הדאטה , כדי שנוכל לבחון את עצמנו בצורה נכונה ואמיתית. את הדטאה של הtrain נשתמש בה בשלב האימון של יצירת הרשת. ואת הדטאה של הtest נשתמש לבדיקה על הרשת שנוצרה ובכך נוכל לדרג את הרשת ולקבוע לה דיוק.

אני בחרתי לחלק את הדאטה 30% test ו- 70% train.

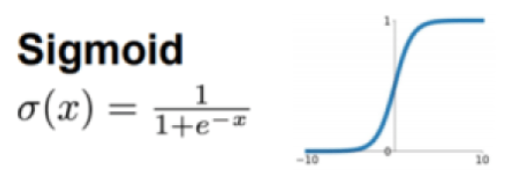
וזה כדי שיהיה לי מספיק מידע בשביל ללמוד ומספיק מידע בשביל לבחון.

**מימוש הרשת:**

את שני הרשתות בניתי בצורה הבאה:

3 שכבות:

* שכבת input: שמכילה 16 נוירונים – שמייצגת את אורך המחרוזת הבינארית הוא 16 ספרות.
* שכבת hidden: שמכילה 10 נוירונים – נבחרה לאחר ניסוי וטעיה, בדיקה על מספר שונה של נויירונים , אך זה נתן לי את הדיוק הטוב ביותר.
* שכבת output: שמכילה בתוכה נוירון אחד – הסיוג האם המחרוזת מתאימה לחוקיות או לא ערכים 1/0.

**פונקצית אקטיבציה:**

פונקציה האקטיבציה שבחרתי לממש היא פונקצית הsigmoid.

זאת לאחר ניסוי וטעיה עם סוגים שונים של פונקציות אקטיבציה שונות לדוגמא relu.

בחרתי בפונקציה זאת כי היא הביא לי את הדיוק הגבוהה ביותר בזמנן הרצוי.

**אלגוריתם גנטי:**

את שלב יצירת הרשת עטפתי באלגוריתם גנטי אשר יעזור לי לשפר ולמצוא את המשקלים הטובים והמתאימים ביותר לרשת שלי כדי לקבל את הסיוג שייתן לי את אחוז הדיוק הגבוה ביותר.

הפרמטרים שהשתמשתי לאלגוריתם הגנטי:

* population\_size – גודל האוכלוסיה – אני בחרתי בגודל של 120. זאת לאחר ניסוי וטעיה, בדיקה של מספר גדלים, בחרתי בגודל זה שהצליח להביא אותי לאחוז דיוק גבוה ולהרצת האלגוריתם במשך זמן סביר.
* generation – אני בחרתי בגודל 60, כי בפחות מזה לא הצלחתי להגיע לאחוז דיוק גבוהה מ 93% ורציתי לנסות להגיע לאחוז דיוק יותר גבוה אך שמבחינת זמן שהאלגוריתם לא ירוץ שעות לכן החלטתי להגביל ל 150 דור.

**החוקיות שמאחורי שני הרשתות:**

רשת 0:

החוקיות היא מספר הביטים הדולקים נמצא בטווח בין 8 ל 12 ואזי הביטוי שייך לחוקיות אחרת אינו שייך לה.

דוגמא:

* בחוקיות :

1010010011110110 1  
1110010000101011 1  
0110101110010111 1

* לא בחוקיות:

1000100010101000 0  
0000101010101001 0  
0000011000110001 0

רשת1:

החוקיות היא מספר הביטים הדולקים קטן מ 8 ואזי הביטוי שייך לחוקיות אחרת אינו שייך לה.

דוגמא:

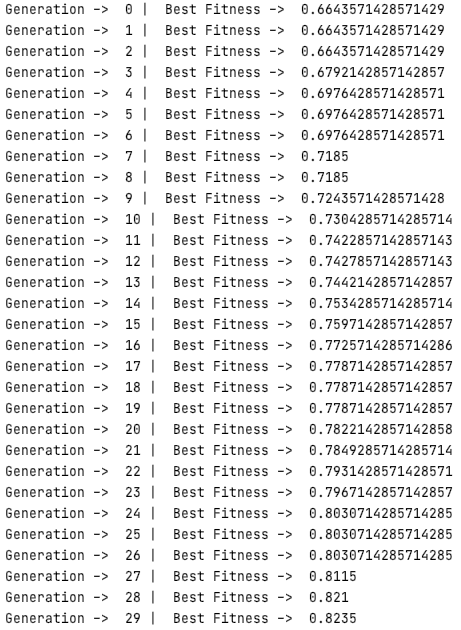
* בחוקיות :

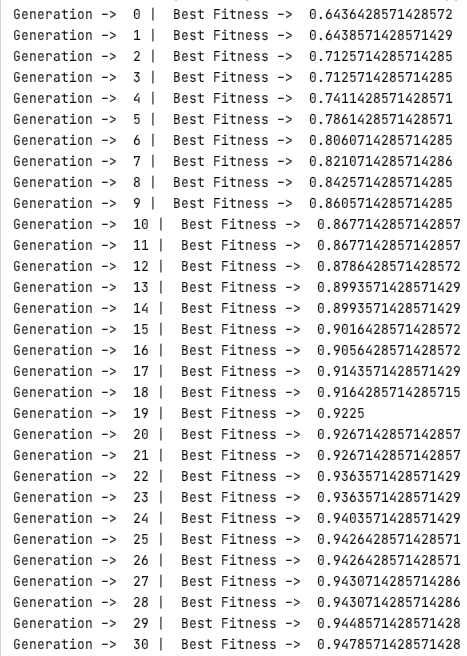
0101101100010100 1  
1101100000010011 1  
1100000000010110 1

* לא בחוקיות:

1101100111100011 0  
1000110000111111 0  
0010001101101111 0

**תצלומים מתהליך האימון:**

רשת 0:



רשת 1:

**אחוזי דיוק:**

רשת 0: 0.90%

רשת 1: 0.97%