**ממן 12 – חישוביות ביולוגית – איילה שעובי מן, ת.ז. 200244242**

# שאלה 1:

1. *שאלה 4, ע״מ 133: תאר/י כיצד למצוא את הערכים של הממזערים את הפונקציה תוך שימוש באלגוריתם גנטי:*

***תיאור האלגוריתם:***

1. *ייצוג: נייצג כל פתרון אפשרי ע״י 2 מחרוזת באורך 5 ביטים. 5 ביטים מספיקים כדי לייצג מספר בין 0 ל-31, טווח הערכים האפשרי בעבור y. וכן בעבור ישנם 31 ערכים אפשרים רצופים ולכן נוכל לבצע הזזה.*
2. *אוכלוסייה ראשונית: בחירה אקראית של n זוגות (x,y) אפשריים*
3. *ערך הכשירות: מידת ההצלחה של פרט i באוכלוסיה יהיה הערך המקסימלי של הפונקציה g באוכלוסייה, פחות ערך הפונקציה בעבור פרט i.*
4. *(זאת לאחר הזזת הייצוג העשרוני בעבור x ב31 יחידות*

*בעבור כל דור G:*

1. *חישוב ערך בעבור כל בעבור i בין 1 ל- n*
2. *אליטיזם- נבחר אחוז מסויים מתוך הפרטים הטובים ביותר, בעלי ערך ה-fitness הגבוהה ביותר, ונעביר אותם לדור הבא, G+1.*
3. *בחירת זוג הורים כאשר ההסתברות לבחירת פרט כהורה היא .*
4. *ביצוע פעולת זיווג בהסתברות ובחירת אחד הצאצאים לדור הבא. בזיווג של הורים ( ו- נייצר שני צאצאים חדשים ו- ונבחר אחד מהם לדור הבא.*
5. *חזרה על שלבים4-6 עד שיש n פריטים לדור הבא.*
6. *בעבור כל פרט i, ובעבור כל גן מתוך , בצע מוטציה (בחירה רנדומלית של ערך אפשרי חדש) בהסתברות*
7. *חזרה על שלבים 5-10*
8. *תנאי עצירה:**עד אשר לא מושג שיפור (הפחתה) בערך במשך n\_max דורות, או עד אשר השינוי בין דור לדור מתחת לסף אשר יקבע מראש.*
9. *שאלה 19, ע״מ 135: הסבר מדוע גישת הנישות מקטינה את הסיכוי להתכנסות מוקדמת.*

*בגישה זו אנו מבצעים חלוקה של האוכלוסייה לתתי אוכלוסיות אשר התפתחותן אינה תלויה זו בזו. ערבוב תתי האוכלוסיות מתאפשר פעם במספר דורות.*

*התכנסות מוקדמת אומנם אפשרית בכל אחת מתתי האוכלוסיות בנפרד, אך התכנסות מוקדם בעבור עלל התוכנית תקרה רק אם כל תתי האוכלוסיית יחוו התכנסות מוקדמת וזה יקרה בהסתברות נמוכה מאוד עקב השונות הרבה בין כל תת אוכלוסייה לאחרת.*

*כמו כן בערבוב אוכלוסיית, אנו מגבירים את השונות הפנימים של כל אוכלוסייה ע״י הכנסת פריטים ״מבחוץ״ ובכך גם מקטינים את הסיכוי להתכנסות מוקדמת של כל תת אוכלוסייה.*

# שאלה 2:

1. **תיאור האלגוריתם הגנטי:**
   1. בחירת נקודת התחלה/מוצא וסיום/מטרה
   2. הגרלת אוכלוסייה התחלתית בגודל N, באורך סביר מנקודת ההתחלה
   3. הגרלת מיקום מכשולים
   4. חישוב ערך ה-  fitness של כלל האוכלוסייה
   5. אליטיזם- בחירת הפרטים הטובים ביותר, בעלי ערך ה-fitness הגבוהה ביותר, מעבירים אותם לדור הבא.
   6. בחירת זוג הורים (כרומוזומים), בהסתברות המחושבת ביחס ישיר ל- fitness של הזוג
   7. ביצוע פעולת זיווג בהסתברות גבוהה ובחירת אחד הצאצאים לדור הבא.
   8. חזרה על שלבים e,f,g עד שיש מספיק פריטים לדור הבא.
   9. חזרה על שלבים d-h עד שמושג מסלול מיטבי (בעל אורך מינימאלי) או עד אשר הגענו למספר דורות מקסימאלי.
2. **תיאור הייצוג**
   1. **מרחב התנועה**

תנועת הרובוט מיוצגת כתזוזה על גבי לוח דו ממדי בגודל 10\*10 או 100\*100.

כל כרומוזום מיוצג ע״י אוסף של צעדים כאשר הצעד הראשון מבוצע מנקודת המקור אשר הוגרלה בעבור קריצת האלגוריתם, וכל צעד עוקב מבוצע על פי הכיוון המיוחס לו מתוך סט כיוונין אפשריים { *U=Up, D=Down, L=Left, R= Right* }.

כמות הגנים בכרומוזום נקבעה באופן יחסי לאורך הלוח והינה פי 2.5 ממנו.

* 1. **אוכלוסייה התחלתית**

האוכלוסייה ההתחלתית מורכבת מאוסף כרומוזומים באורך שווה, כאשר כל גן בכרומוזום נבחר באופן רנדומלי מתוך סט הערכים האפשריים.

כמות הכרומוזומים הינה 20, 40 או 60 כרומוזומים (תחת ניסויים שונים).

* 1. **אופרטורים גנטיים** 
     + **אליטיזים**- 10% מהכרומוזומים מהדור הנוכחי מועברים לדור הבא ללא כל שינוי
     + בחירת הורים לפעולת שיחלוף - בשיטת הרולטה, בחירה של זוג כרומוזומים מדור נוכחי בהסתברות  P\_c.
     + **שחלוף**- בהסתברות 0.7, יבוצע שחלוף בין זוג כרומוזומים אשר נבחרו כהורים, בנקודה x= אינדקס רנדומלי עד אורך הכרומוזום אחד משני הצאצאים שנוצרו, בבחירה רנדומלית, יועבר לדור הבא.

בהסתברות 0.3 לא יבוצע שיחלוף, ובבחירה רנדומלית, אחד מזוג הכרומוזומים אשר נבחרו כהורים, יועבר לדור הבא ללא שינוי.

* + - **מוטציה**- בעבור כל כרומוזום, וכל גן, תבוצע מוטציה (הגרלה של ערך חדש), בהסתברות 0.15.
    - **Fitness**- לצורך חישוב הכשירות של כרומוזום בחרתי להתחשב ב-2 פרמטרים
      * 1. מחיר המסלול – זהו אורך המסלול שמבצע הכרומוזום. נעדיף מסלולים קצרים יותר
        2. מרחק מהיעד הסופי של הרובוט- אם הכרומוזום מסיים את מסלולו בנקודה שונה מיעד הרובוט, נוסיף עונש השווה למרחק מנהטן מיעד הרובוט
        3. על כל מכשול שדרכנו עליו, נוסיף עונש קבוע של 10
        4. על מנת להפוך את המחיר המתקבל לפונקציית כשירות, נחזיר את המחיר בחזקת מינוס 1, כך, בעבור מחיר גבוה נקבל כשירות נמוכה ובעבור מחיר נמוך נקבל כשירות גבוהה כנדרש.
  1. **תנאי עצירה**

הארגוריתם יעצור כאשר אחת משתי האפשרויות הבאות יתממשו:

* + - הגעת האלגוריתם למספר מקסימאלי של דורות
      * בעבור לוח של 10\*10, קבעתי את מספר הדורות המקסימאלי להיות 200
      * בעבור לוח של 100\*100, מספר הדורות המ׳סימאלי הינו 500
    1. נמצא מסלול אופטימאלי בדור הנוכחי- שיעדו שווה ליעד הרובוט ואורכו שווה למרחק מנהטן בין נקודות המקור לנקודת היעד
  1. **מכשולים**

ניתן להגדיר ריצה עם 0%, 10%, 20% או 30% מכשולים על גבי הלוח

* 1. **מידע נוסף**
     + אורך מסלול של כרומוזום- המידה ממסלול הכרומוזום חרג מגבולות הלוח או נתקע במכשול, אורך המסלול יקבע לפי הנקודה האחרונה שהגיע אליה על הלוח. במידה והכרומוזום הגיע ליעד הרובוט, אורך המסלול יקבע לפי מספר הצעדים שבוצעו עד להגעה לנקודת היעד
     + בחירת נקודות מקור ויעד- נקבע רנדומאלית מבין כלל משבצות הלוח

1. **ריצה**

מצורף קובף batch בשם genetic\_alg.bat להרצת התוכנית.

מימוש האלגוריתם בוצע ע״י שלושה סקריפטים:

* + - Genetic\_alg.py
    - Grid.py
    - Main.py

בעבור כל הרצה ניתן לבחור את ערכי הפרמטרים הבאים:

* + - גודל הלוח (10\*10 או 100\*100)
    - גודל האוכלוסייה ( מתוך 20, 40 או 60)
    - אחוז המכשולים (0-30%)

סה״כ בחנתי 12 שילובים שונים של הפרמטרים הנ״ל המפורטים בסעיף הבא

1. **חקירת האלגוריתם הגנטי תחת שילוב פרמטרים שונים והצגת תוצאות**

**פירוט ההרצות שבוצעו:**

**פירוט הניסויים:**

|  |  |  |  |
| --- | --- | --- | --- |
| **מזהה ניסוי** | **לוח** | **אוכלוסייה התחלתית** | **מכשולים** |
| 1 | 10\*10 | 20 כרומוזומים | 0% (ללא) |
| 2 | 10\*10 | 40 כרומוזומים | 0% (ללא) |
| 3 | 10\*10 | 60 כרומוזומים | 0% (ללא) |
| 4 | 10\*10 | 60 כרומוזומים | 10% |
| 5 | 10\*10 | 60 כרומוזומים | 20% |
| 6 | 10\*10 | 60 כרומוזומים | 30% |
| 7 | 100\*100 | 20 כרומוזומים | 0% (ללא) |
| 8 | 100\*100 | 40 כרומוזומים | 0% (ללא) |
| 9 | 100\*100 | 60 כרומוזומים | 0% (ללא) |
| 10 | 100\*100 | 60 כרומוזומים | 10% |
| 11 | 100\*100 | 60 כרומוזומים | 20% |
| 12 | 100\*100 | 60 כרומוזומים | 30% |

**הצגת התוצאות:**

1. ניסויים 1-3- לוח 10\*10, ללא מכשולים

|  |  |  |
| --- | --- | --- |
| ניסוי 1- לוח סופי | ניסוי 2- לוח סופי | ניסוי 3- לוח סופי |
| ex1 | ex2 | ex3 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ניסוי | מסלול מינימאלי לאורך הריצה | פיטנס מקסימאלי בדור אחרון | פיטנס מקסימאלי לאורך הריצה | דור בו הושג פיטנס מקסימרלי | האם הושג מסלול אופטימאלי | מספר דורות עד לפתרון | טרנד כשירות |
| **1**  הושג פתרון אופטימאלי המגיע ליעד הסופי באורך 9, והתקבל לאחר 32 דורות | 9 | 0.08 | 0.12 | 9 | כן | 32 |  |
| 2  הושג פתרון אופטימאלי המגיע ליעד הסופי באורך 10, והתקבל לאחר 19 דורות | 10 | 0.055 | 0.057 | 15 | כן | 19 |  |
| 3  הושג פתרון אופטימאלי המגיע ליעד הסופי באורך 7, והתקבל לאחר 17 דורות | 7 | 0.29 | 0.38 | 1 | כן | 17 |  |

1. ניסויים 4-6- לוח 10\*10, עם מכשולים

|  |  |  |
| --- | --- | --- |
| ניסוי 4- לוח סופי | ניסוי 5- לוח סופי | ניסוי 6- לוח סופי |
| ex4 | ex5 | ex6 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ניסוי | מסלול מינימאלי | פיטנס מקסימאלי בדור אחרון | פיטנס מקסימאלי לאורך הריצה | דור בו הושג פיטנס מקסימרלי | האם הושג מסלול אופטימאלי | מספר דורות עד לפתרון | טרנד כשירות |
| 4  הושג פתרון אופטימאלי המגיע ליעד הסופי באורך 8, והתקבל לאחר 22 דורות | 8 | 0.29 | 0.38 | 10 | כן | 22 |  |
| 5 | 9 | 0.065 | 0.065 | 42 | כן | 42 |  |
| 6 | 12 | 0.025 | 0.055 | 5 | לא | 301 |  |

1. ניסויים 7-9- לוח 100\*100, ללא מכשולים

|  |  |
| --- | --- |
| ניסוי 7- לוח סופי | ex7 |
| ניסוי 8- לוח סופי |  |
| ניסוי 9- לוח סופי | ex9 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ניסוי | מסלול מינימאלי | פיטנס מקסימאלי בדור אחרון | פיטנס מקסימאלי לאורך הריצה | דור בו הושג פיטנס מקסימרלי | האם הושג מסלול אופטימאלי | מספר דורות עד לפתרון | טרנד כשירות |
| 7 | 51 | 0.1 | 0.15 | 177 | כן | 253 |  |
| 8 | 34 | 0.09 | 0.1 | 208 | לא | 222 |  |
| 9 | 121 | 0.0315 | 0.033 | 490 | לא | 500 |  |

1. ניסויים 10-12- לוח 100\*100, עם מכשולים

|  |  |
| --- | --- |
| ניסוי 10- לוח סופי | ex10 |
| ניסוי 11- לוח סופי | ex11 |
| ניסוי 12- לוח סופי | ex12 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ניסוי | מסלול מינימאלי | פיטנס מקסימאלי בדור אחרון | פיטנס מקסימאלי לאורך הריצה | דור בו הושג פיטנס מקסימרלי | האם הושג מסלול אופטימאלי | מספר דורות עד לפתרון | טרנד כשירות |
| 10 | 250 | 0.3 | 0.033 | 177 | לא | 500 |  |
| 11 | 421 | 0.027 | 0.029 | 497 | לא | 500 |  |
| 12 | 475 | 0.0253 | 0.0253 | 500 | לא | 500 |  |

1. **ניתוח התוצאות והסקת מסקנות**
   1. **בחינת התוצאות בעבור לוח 10\*10:**

ניתן לראות כי אנו מגיעים לפתרון האופטימלי באופן מהיר, וללא ניצול של מספר הדורות המקסימאלי.

כמו כן, מספר הדורות הנדרש להתכנסות **קטן כאשר אנו מגדילים את האוכלוסיה ההתחלתית** מ-20 ל-60.

עם זאת, כאשר מדובר על מספר דורות נמוך ולוח קטן, לא ניתן לראות הוכחה לשיפור מגמתי בערך הכשירות המיטבי בכל דור.

**כאשר מוסיפים מכשולים ללוח**, ניתן לראות כי ככל שכמות המכשולים על הלוח גדלה, כך יש צורך ביותר דורות על מנת להגיע אל היעד אשר הוגרל. בסט הניסויים הרלוונטים (4-6) אכן ניתן לראות מגמת שיפור ברורה בערך הפיטנס לאורך הדורות, עד להגעה ליעד.

מסקנה הרלוונטית למימוש נוסחאת הכשירות היא כי יש צורך במתן עונש בעבור חזרתיות על תאים לאורך המסלול עד להגעה ליעד, לשם מניעת מעגלים. כך לדוגמא, המסלול המתקבל בניסוי 6, התקבל לאחר 60 דורות, אך אינו אופטימלי, ומתן עונש כנ״ל היה יכול להוביל לשיפורו לאורך הדורות 61-300.

* 1. **בחינת התוצאות בעבור לוח 100\*100:**

ניתן לראות כי בעבור לוח גדול יותר, כאשר מעלים את האוכלוסייה ההתחלתית ל-60, בהשוואה ל-20 או 40, ההתכנסות הופכת לקשה יותר.

ההסבר ההגיוני לכך הוא כתוצאה מלוגיקת הזיווג. זוג ההורים נבחרים מהאוכלוסייה הכוללת של הדור הנוכחי, ולכן כאשר יש יותר אפשרויות זיווג, תיתכן שונות גבוהה יותר בין הפריטים שתוביל קפיצות בין מסלולים אפשריים במקום התכנסות, שיפור של פונקציית הכשירות תעלה את ההסתברות לבחירת הורים בעלי כשירות גבוהה ותוריד את ההתברות לבחירת הורים בעלי כשירות נמוכה באופן קיצוני יותר.

ניתן לבחון זאת ע״י הרצה עם גדולה באופן משמעותי, לדוגמא בגודל 100.

**כאשר מוסיפים מכשולים**, בניסויים 10-12, ניתן לראות כי האלגוריתם לא הצליח להגיע ליעד באף אחד מהריצות, למרות שיפור ניכר מדור לדור.

בעבור לוח עם 30% מכשולים, לאורך הריצה ניתן היה להבחין בשינויים דחופים של המסלול באופן שאינו מקדם את הרובוט לקראת הגעה ליעד. הסיבה לכך היא לולאות במסלולים כך שמסלול שקוצר עקב הגעה למכשול, יכול לקבל מחיר שקול ואפילו נמוך יותר ממסלול שמיצה את כלל צעדיו ללא התקלות במכשול כאשר בפועל, יש לשנות את פונקציית המחיר כך שהשני מבין השניים יועדף ויהיה בעל כשירות גבוהה יותר. זאת בהתאמה להתנהגות שהשתקפה בניסוי 6.

עם זאת, גם בהתכנסות לפתרון שאינו אופטימלי, והתקעות האלגוריתם, ע״י הכנסת רכיב המוטציה בהסתברות גבוהה יחסית, ניתן היה לראות כי לאחר מספר דורות האלגוריתם יכל היה לצאת מהמצב ולהמשיך לשפר עמדות. בעקבות זאת, אני מאמינה כי ע״י הגדלת מספר הדורות המקסימאלי ניתן היה להגיע ליעד הרובוט במרבית המקרים.

* 1. **מסקנות:**
     + הגדלת האוכלוסייה ההתחלתית משפרת את מהירות ההתכנסות אך עם זאת בעבור לוח קטן, אוכלוסייה התחלתית קטנה היא מספקת כך שניתן לקחת אוכלוסייה ביחס לגודל הלוח
     + נדרשת התאמה ללוגיקת הזיווג בשביל למנוע בחירה של הורים בעלי כשירות נמוכה בהסתברות גבוהה יחסית
     + שיפורים נוספים הנדרשים לפונקציית הכשירות: עונש על מעגלים במסלול, עונש על סיום מסלול שלא כתוצאה מהגעה ליעד
     + הרצה בעבור מספר דורות גבוהה יותר יכולה להוסיף ולשפר את ערך הכשירות, כאשר אכן ישנו שיפור מדור לדור.
     + הסתברות למוטציה מונעת מצבים של התכנסות לפתרון לא אופטימלי ויציאה ממצב קפאון