# תרגיל 1 – אלגוריתמים בביולוגיה חישובית

**שאלה 1 – מספר העימודים האפשריים**

יהיו  *שני רצפי באורך כל אחד.*

*נרצה להראות כי מספר העימודים האפשריים של מול הוא אקספוננציאלי ב- .*

*נבחין כי הגודל המקסימלי של כלל הסידור יהיה .*

*נפשט את הבעיה (נצמצם אפשרויות סידור) ונגיד בלי הגבלת הכלליות שאנחנו בוחרים מיקומים של הרצף ומיקומי הבסיסים של הרצף יוגדרו ככה שהם בוחרים מיקום אקראי בודד בין לבין .*

*לכן, הבעיה שקולה לבחירת מקומות להעמדה של הבסיסים של מתוך . וזו בעיה מוכרת בקומבינטוריקה – לבחירת מקומות להעמדת הבסיסים של מתוך מיקומים קיימות אפשרויות (עם שמירה על סדר רצף ה-).*

*נראה כי כבר הביטוי לבחירת מיקומי בסיסים של מהרצפים הוא אקספוננציאלי ב- .*

*לכן, נקבל כי מספר העימודים האפשריים של מול הוא לפחות אקספוננציאלי ב- .*

***שאלה 2:***

***ניתוח זמן ריצה***

*נסמן את אורך סדרת הנקודות – , קנס על כל סגמנט - , אורך מקסימלי של סגמנט - .*

*A math equation with black text

Description automatically generated with medium confidenceמטרת האלגוריתם למצוא את הארגומנט המינימלי של נוסחת ה- (סכום הפרשי הריבועים) על מנת להגיע למזעור המרחקים של כל נקודה איבר בסדרה שנקבל כקלט) לערך הממוצע בסגמנט אליה נשייך, בתוספת קנס p :*

*אופן פעולת האלגוריתם הינו מעבר על כל הנקודות בסדרה (קיימות כאלה כמספר האיברים), ועבור כל נקודה בודקים מה הסגמנט האידיאלי (שייתן מינימלי) שנגמר באותה נקודה – כלומר נקודת התחלה אידיאלית, בודקים עבור כל נקודה את ה- עבור כל מקטע שמתחיל עד נקודות לפניה:*

*כאשר היא נקודת ההתחלה של הסגמנט, היא נקודת הסוף ו- הערך הממוצע של הסגמנט.*

*עבור כל נקודה עדכנו את נקודת ההתחלה ואת מערך ה- במידה ומצאנו מינימום חדש.*

*לכן, נקבל כי אנחנו מבצעים חישובים של ה- , כל חישוב על פי הנוסחה הוא .*

*בנוסף, לחישוב ה של הסגמנט בצורה יעילה השתמשנו בשני מערכי עזר:*

*האחד לחישוב סכום מצטבר –*

*השני לחישוב סכום ריבועי מצטבר –*

*כך שבכל פעם יכולנו לחשב את הממוצע של הסגמנט ואת הנוסחה של במספר פעולות קבוע (חמש).*

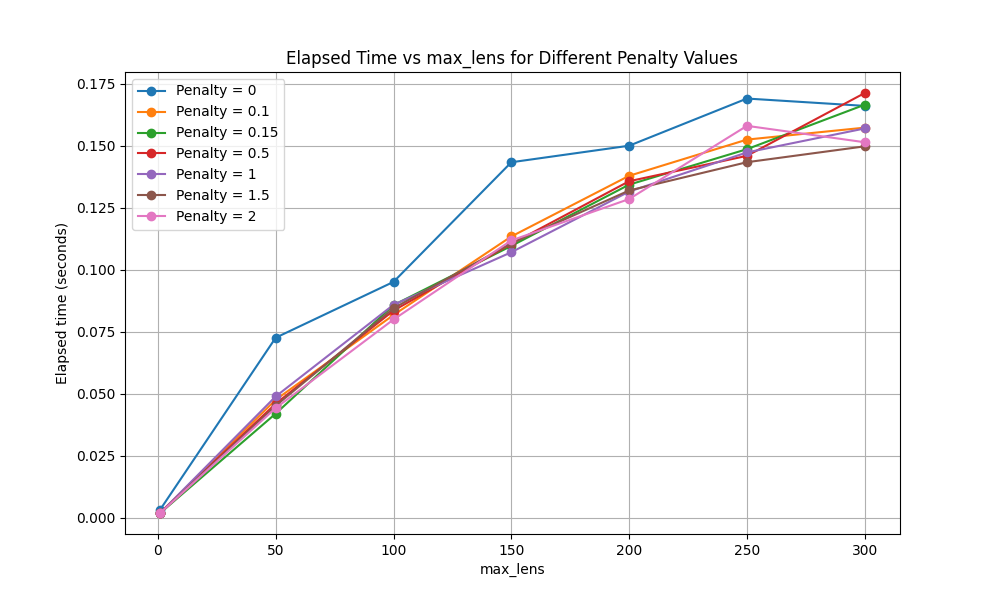
*סך הכל לאלגוריתם הדינאמי -*

*לאחר סיום החלק הזה באלגוריתם אנחנו משחזרים את הפתרון האופטימלי שנמצא על ידי מעבר על איברים ממערך שיצרנו לשם הפתרון – נעבור על כל איבר ב- שייצג עבור כל יציין איפה הסגמנט מתחיל. למצוא סגמנטים כאלו.*

***זמן הריצה הכולל:***

*במקרה הגרוע ביותר נעבור על n נקודות לכן שלב זה מתבצע ב- כי לא ייתכנו יותר סגמנטים ממספר הנקודות ( ).*

***לכן נקבל סה"כ שסיבוכיות זמן הריצה של האלגוריתם הדינאמי הוא .***

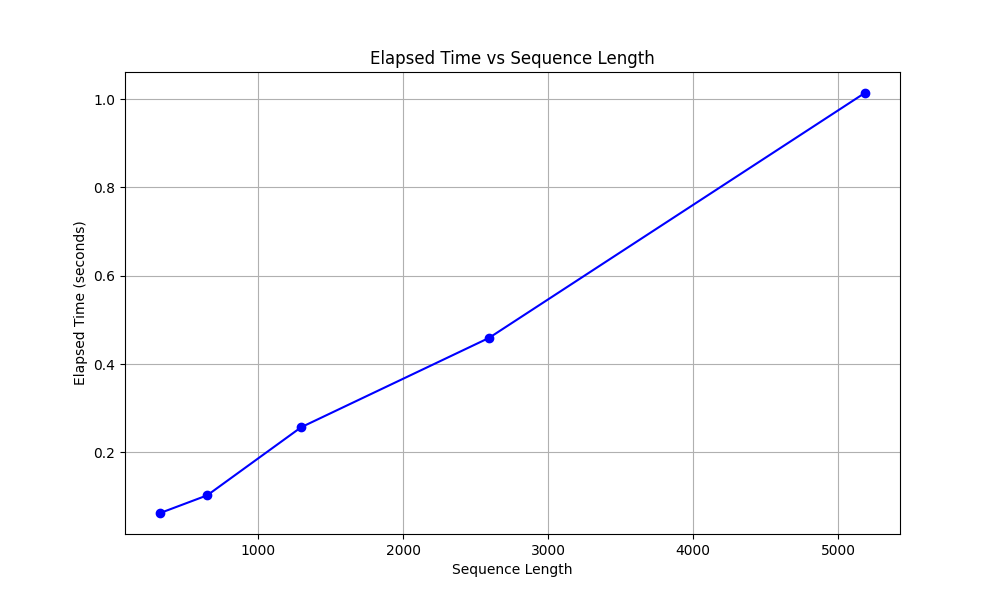
*נרצה להראות את השינוי בזמן הריצה כתלות ב- (אורך הסגמנט המקסימלי):*

*יצרנו 7 קווים (לכל ערך שונה) המציגים את זמן הריצה כתלות בערכי שונים. ניתן לראות כי אכן רואים קשר כמעט לינארי בין הגבלת אורך הסגמנט המקסימלי לבין זמן ריצה של התוכנית – ככל שהגבלת הגודל גדולה יותר זמן הריצה עולה.*

*A graph of different colored lines

Description automatically generatedנרצה לוודא כי אכן גודל הקנס לא משפיע על זמן הריצה:*

*כפי שציפינו, ניתן לראות כי לא ניתן להצביע על קשר בין גודל הקנס לבין זמן ריצת התוכנית.*

*נראה כי אכן קיים קשר לינארי בין אורך סדרת הנקודות לבין זמן הריצה של התוכנית כאשר קבועים:*

**A graph with numbers and points

Description automatically generated with medium confidence***כעת, נביט בגרפים שונים על ידי שינוי הפרמטרים עם קובץ ה- שסופק:*

**A graph with numbers and points

Description automatically generated with medium confidence**

*מהשוואה בין הגרפים ניתן לראות כי בשל העלאת ערך ה- בריצה השנייה נוצרו פחות סגמנטים וכתוצאה מכך הסגמנטים ארוכים יותר (וכן כי אפשרנו אורך ארוך יותר לכל סגמנט), וכן נוצר סגמנט ארוך שכולל יותר מ-50 נקודות (154 עד 214).*

***בחינת מקרי קיצון:***

*A graph with blue and green lines

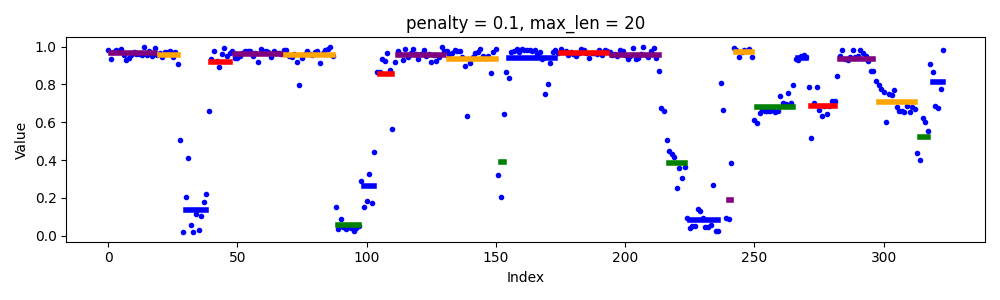
Description automatically generated****A graph with blue dots

Description automatically generatedA graph with numbers and points

Description automatically generated with medium confidenceA graph with blue and red dots

Description automatically generatedללא הגבלת עם הולך וגדל:***

*כפי שניתן לצפות, רואים שככל שנעלה את הקנס על הסגמנטים נקבל פחות ופחות סגמנטים.*

***A graph with numbers and colored dots

Description automatically generated with medium confidencepenalty קבוע עם משתנה:***

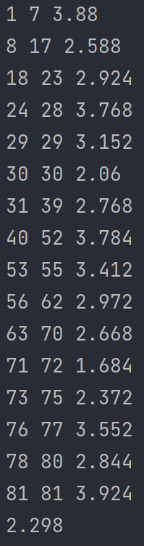
***A graph with numbers and points

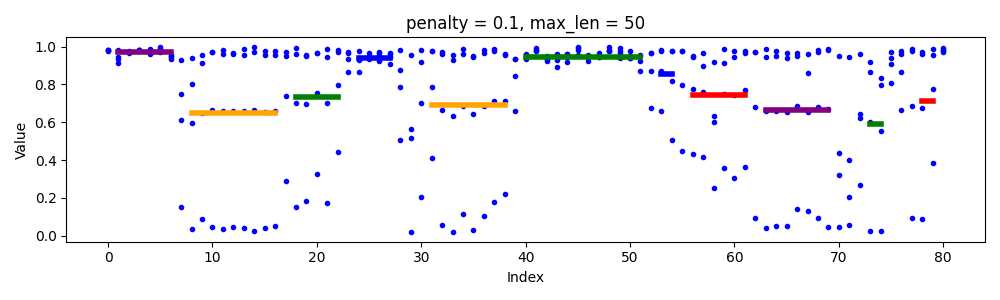
Description automatically generated with medium confidence****A graph with numbers and points

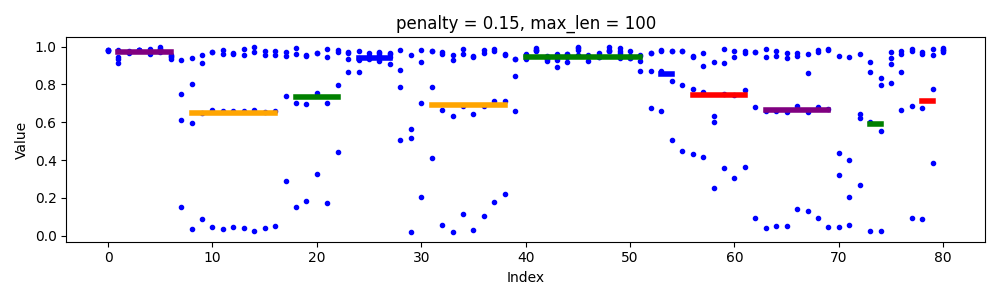
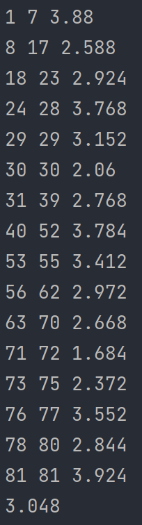
Description automatically generated with medium confidence*

*מעניין לראות מרצף הגרפים כי ככל שאנחנו מאפשרים אורך מקסימלי גדול יותר (10,20,60) נראה כי אכן נקבל סגמנטים באורכים גדלים, אומנם עד לשלב מסוים שבו התוכנית תגיע למינימום של ערך הcost- עבור חלוקה לסגמנטים מסוימת והחל משלב זה העלאת אורך הסגמנט המקסימלי לא תשפיע על החלוקה לסגמנטים (כפי שרואים בהשוואה בין ו- ).*

***חלק בונוס – סגמנטציה למידע רב ערוצי:***

*נריץ את קובץ הקלט של חלק הבונוס עם ערכים שונים של ו- :*

**

**

*בגרפים שמנו את כל הנקודות מכל הערוצים על הגרף וניתן לראות כי אכן הסגמנטים הנבחרים מצאו איזון כלשהו בין כל הערוצים השונים. ואף ניתן לראות כי גם ערך ה-cost גבוה יותר מערך ה- כאשר חישבנו בערוץ בודד בחלק הקודם.*