

## דף תרגילים אחד-עשר ביסודות מדעי המחשב

נתון קובץ טקסט. קובץ הטקסט מורכב משורות ובכל שורה כתובים נתונים עבור דוגמה מסויימת. כל הדוגמאות הן מופעים של אותה מחלקה.

1. כתבו פונקציה שמעבירה את כל העמודות לרשימות. כל הרשימות מלבד האחרונה צריכות לעבור כיוול. ניתן להניח שיש סה"כ 5 רשימות.
2. צרו מחלקה עבור הדוגמאות בקובץ. לכל מופע במחלקה צריך להיות שם, תכונת-נתון אחת שהיא וקטור (מערך), ותכונת נתון שנייה שהיא מספר. במחלקה צריכה להיות פונקציה לחישוב המרחק בין שני מופעים, שהוא המרחק האוקלידי בין הוקטורים המשוויכים להם, ופונקציה המחשבת את מימד הוקטורים.
3. כיתבו פונקציה שיוצרת מהרשימות הראשונות (מלבד האחרונה) וקטורי תכונות. כל וקטור מיוצג כמערך ע"י `pylab.array`. הפונקציה יוצרת מכל וקטור ומספר מהרשימה החמישית איבר במחלקה מתרגיל 1, ומקבצת את האיברים לרשימה.
4. צרו מחלקה עבור צביר של דוגמאות, כל דוגמה היא מסוג המחלקה מתרגיל 1. לכל מופע במחלקה צריכים להיות שני נתונים: הראשון הוא אוסף דוגמאות והשני הוא המרכז שלהן. הוסיפו למחלקה את הפונקציות הבאות:
  - א. פונקציה שמחשבת את מרכז הצביר.
  - ב. פונקציה שמחזירה את ההפרש בין מרכז הצביר למרכז של אוסף דוגמאות אחר, ומשייכת את אוסף הדוגמאות החדש לצביר במקום אוסף הדוגמאות הקודם.
  - ג. פונקציה שמחזירה את ה `variability` של צביר.
5. כיתבו פונקציה שמקבלת רשימת דוגמאות ומספר  $k$  חיובי כלשהו, ומחזירה  $k$  צבירים, לפי השלבים הבאים:
  - א. הפונקציה בוחרת בצורה אקראית  $k$  דוגמאות מתוך רשימת הדוגמאות שהיא מקבלת, ויוצרת מהן רשימת צבירים התחלתית.
  - ב. הפונקציה עוברת על רשימת הדוגמאות ויוצרת רשימת צבירים חדשה. כל איבר במיקום  $i$  ברשימה החדשה הוא צביר שמכיל את כל הדוגמאות שמרכז הצביר ה  $i$  הכי קרוב אליהן.הפונקציה חוזרת על שלב ב אם מרכז הצביר  $i$  שונה ממרכז הצביר במקום ה  $i$  ברשימה החדשה. שימו לב, הפונקציה שבודרת אם המרכז שונה גם מחליפה את הדוגמאות לדוגמאות החדשות.
6. כתבו פונקציה שמחשבת את ה `dissimilarity` עבור רשימת הצבירים.
7. עבור הנתונים בקובץ הטקסט הצורף  $k=4$ , הדפיסו את ה `dissimilarity` והדפיסו עבור כל צביר כמה דוגמאות יש בו, כמה מתוכן קיבלו את התווית 0 ומהו מרכזו.

בהצלחה!