

# מבוא לבינה מלאכותית

## תרגיל בית 3

### הבעיה

ברצוננו לחלק קבוצת נקודות במרחב דו-מימדי לclusters באמצעות Hierarchical clustering. כל נקודה מיוצגת ע"י זוג מספרים  $\langle x, y \rangle$  המתאר את מיקומה במרחב. בתרגיל זה נעמוד על ההבדלים בין שיטות חישוב המרחק הבאות: average link ו single link.

### תזכורת

single link: distance between two clusters is the shortest distance between a pair of elements from the two clusters.

average link: distance between two clusters is the average distance between each pair of elements from the two clusters.

### מימוש

**קלט:** התוכנית תקרא את אוסף הנקודות כקלט מקובץ input.txt יחיד. בשורה הראשונה לקובץ יש לכתוב את סוג האלגוריתם חישוב המרחק בו נשתמש; אנו נתמוך ב2 בלבד – average link ו single link. בשורה השניה יהיה כתוב את מספר ה clusters הנדרש. לאחר מכן יכתבו הנקודות באופן הבא:

x1,y1  
x2,y2  
...  
xn,yn

**פלט:** יכתב לקובץ output.txt והוא יכיל את הסיווג לclusters של הנקודות מקובץ הקלט. שים לב, שמות clustering הם מספרים מ1 ועד כמות ה clusters בנדרשת בשאלה.

כדי לקבל פלט אחיד לכל הסטודנטים יש לדאוג שהcluster בו תשובץ הנקודה הראשונה בקובץ הקלט יהיה קלאסטר מספר 1, הנקודה הבאה שאיננה שייכת לקלאסטר 1 תחשב כחלק מקלאסטר 2 וכך הלאה.

בנוסף, יש להגיש קובץ details.txt בו יש לכתוב את שם המגיש באותיות אנגליות קטנות בשורה הראשונה ובשורה השניה את מספר ת.ז.

מצורף קובץ קלט ופלט לדוגמא.

מצורף קובץ details לדוגמא.

יש לתעד את הקוד.

תאריך ההגשה הינו סופי.

בהצלחה,  
אריאל.