

תרגיל זה בנוי על שימוש במערכת compact ובכמה מערכי נתונים שנמצאים על האתר של הקורס <http://horn.tau.ac.il/NCcourse.html>

1. מערך הנתונים של GOLUB מתיחס לתוצאות גנטיות (התבטאויות של תוצרי RNA) ארבע קבוצות של חולי לוקמיה של מצא את ארבע הקבוצות על ידי בצוע clustering על פני 72 החולים בהינתן מטריצת ההתבטאות של 7129 הגנים.

נסה כמה שיטות וכמה פרמטרים. איזה שיטות ופרמטרים נתנו תוצאות טובות?

אם נגדיר את טיב השיטה רק ביחס לחלוקה לקבוצות הגדולות (AML לעומת ALL) האם ניתן לקבל ערכים טובים יותר עבור קריטריון Jaccard? האם צריך לשנות את העיבוד או את השיטה או את הפרמטרים לצורך זה?

2. מערך הנתונים של SPELLMAN מתיחס לתוצאות גנטיות שמתארות את מעגל התחלקות התא של שמר. מערך זה מתאר התבטאות של 798 גנים כפונקציה של 72 נקודות זמן. מטרת התרגיל הזה היא אשכול של כל הגנים. האם ניתן להגיע לחמש הקבוצות השונות האחראיות על הפזות השונות של חלוקת התא?

נסה כמה שיטות וכמה פרמטרים. איזה שיטות ופרמטרים נתנו תוצאות טובות?

3. הגדר אלגוריתם היררכי והפעל אותו על הנתונים הנ"ל. הצע דרך להכללתו במסגרת compact.

שאלות להבהרת נושאים אלה ניתן להפנות אל רועי ורשבסקי  
[royke@post.tau.ac.il](mailto:royke@post.tau.ac.il)