**דוח רידוד מימדים, שיטות אישכול שונות ובדיקות סטטיסטיות**

מגישים: דוד ליפשיץ, עינב שנולד.

הדאטה: ספרות כתב יד, לקחנו רק את התמונות עם התיוגים 5,6,9. סה"כ עבדנו על 543 תמונות, 64 פיצ'רים עבור כל תמונה

תיאור תהליך השוואה בין מודלים:

הרצנו כל מדד כמותי על 10 סאבסטים של הדאטה (דגמנו כל פעם 50% מהדאטה עם רנדום סיד שונה)

השוואה בין הייפר פרמטרים (עבור כל הרצה של אלג' אישכול ואלג' רידוד מימדים):

* השוואה מול מדד צללית\*
* במקרה והיו יותר מ2 קונפיגורציות של הייפר פרמטרים לקחנו את 2 הקונפיגורציות שממוצע מדד הצללית עליהם היה מקסימלי
* הפעלת paired t-test –
  + אם היפותזת ה0 נדחתה לקחנו את הקונפיגורציה עם ממוצע ציוני הצללית הגבוה ביותר
  + אם ההיפותזה התקבלה – לקחנו את הקונפיגורציה הראשונה שבחנו עליה

השוואה בין אלגוריתמי רידוד מימדים עבור כל אלגוריתם אישכול:

* השוואה מול מדד מידע משותף (mutual information)
* לקחנו את שני אלגוריתמי הרידוד שקיבלו את הממוצע (על המדד המשותף) הגבוה ביותר.
* הפעלת paired t-test –
  + אם היפותזת ה0 נדחתה לקחנו את אלגוריתם הרידוד עם ממוצע ציוני המידע המשותף הגבוה ביותר
  + אם ההיפותזה התקבלה – לקחנו את אלגוריתם הרידוד הראשון שבחנו עליו

השוואה בין אלגוריתמי אישכול:

* השוואה מול מדד מידע משותף (mutual information)
* הפעלת anova test
  + אם היפותזת ה0 נדחתה לקחנו את אלגוריתם האישכול עם ממוצע ציוני המידע המשותף הגבוה ביותר
  + אם ההיפותזה התקבלה – לקחנו את אלגוריתם האישכול הראשון שבחנו עליו

\*עבור קונפיגורציות שאלג' האישכול החזיר מחלקה יחידה, שמנו במדד הצללית ערך שלילי גדול (10-) כיוון שמדד הצללית לא עובד במקרים כאלה ואם קיבלנו רק מחלקה יחידה ביצועי האלגוריתם מבחינתנו גרועים ביותר

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

Kmeans parameters:

N\_clusters = 3

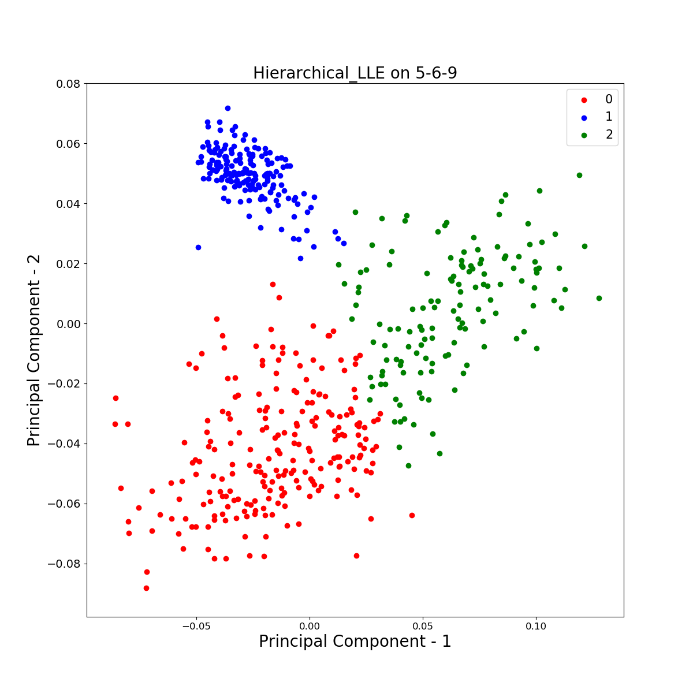
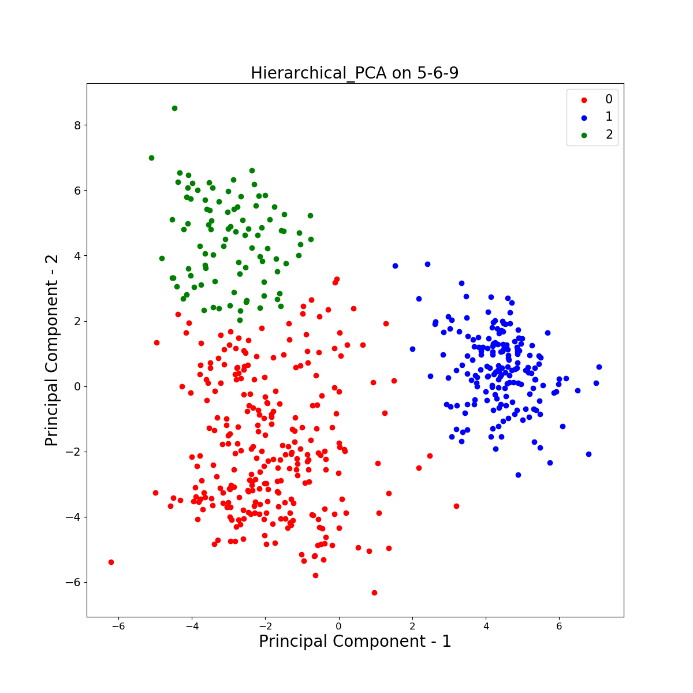
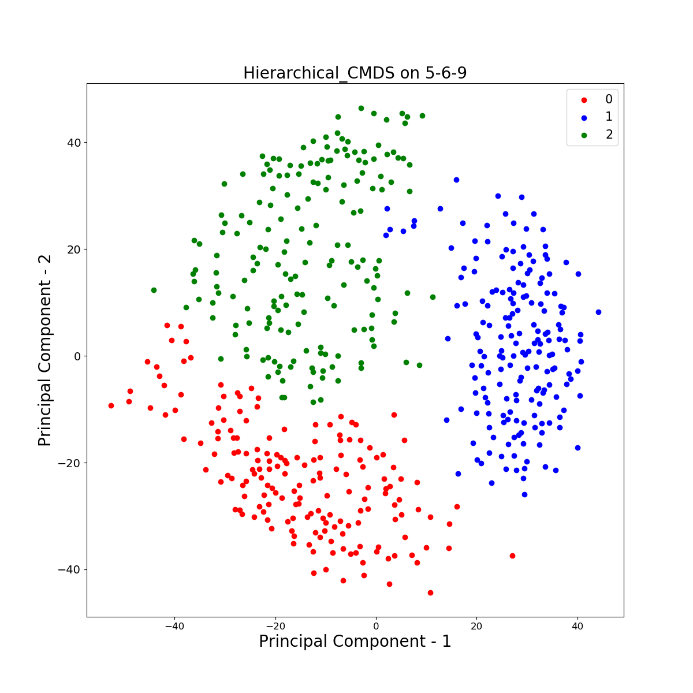
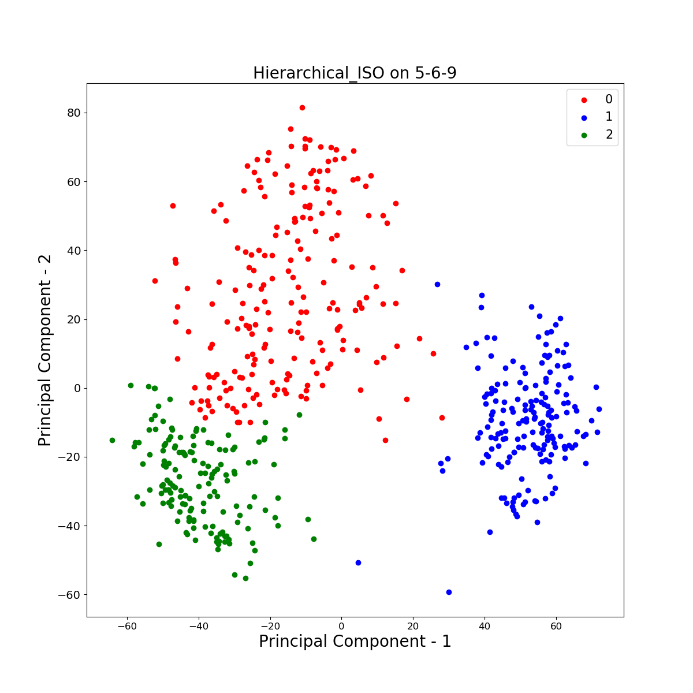
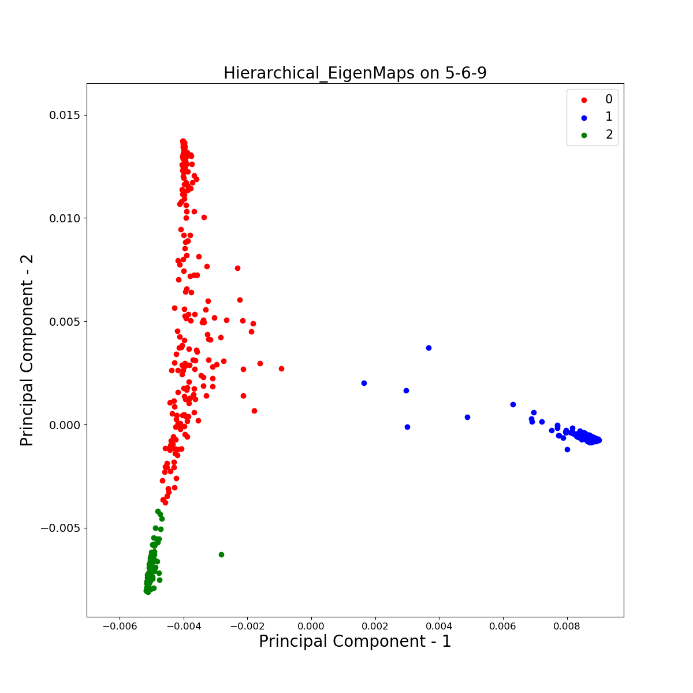
השוואה בין הייפר פרמטרים:

עבור kmeans בדקנו האם עדיף לבצע אישכול על הדאטה לפני או אחרי ביצוע רידוד המימדים

השוואה בין אלגוריתמי רידוד:

בדיקת השפעה של הרצת KMEANS בקונפיגורציות שונות על מדד צללית (עבור כל אלגוריתם רידוד מימדים בדקנו האם יש השפעה לכך שמריצים את אלג' האישכול לפני ביצוע רידוד מימדים או אחרי)





Hierarchical clustering parameters:

N\_clusters = 3

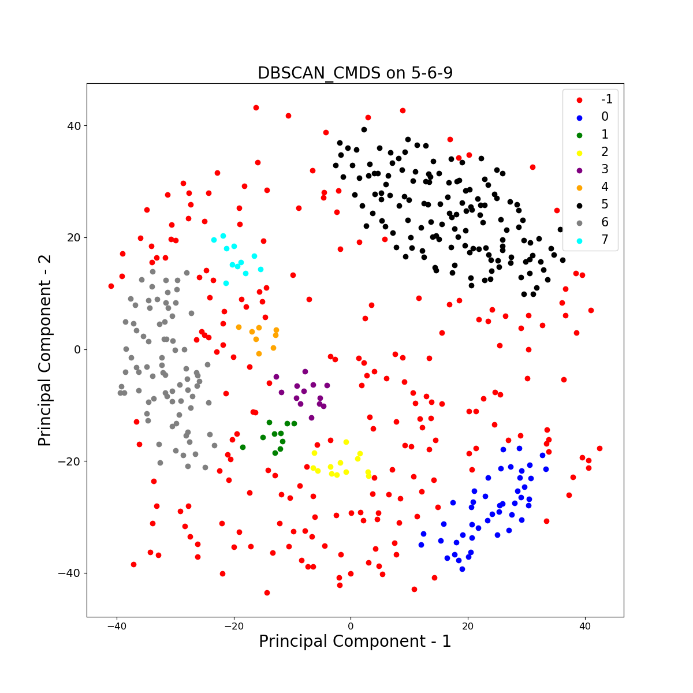
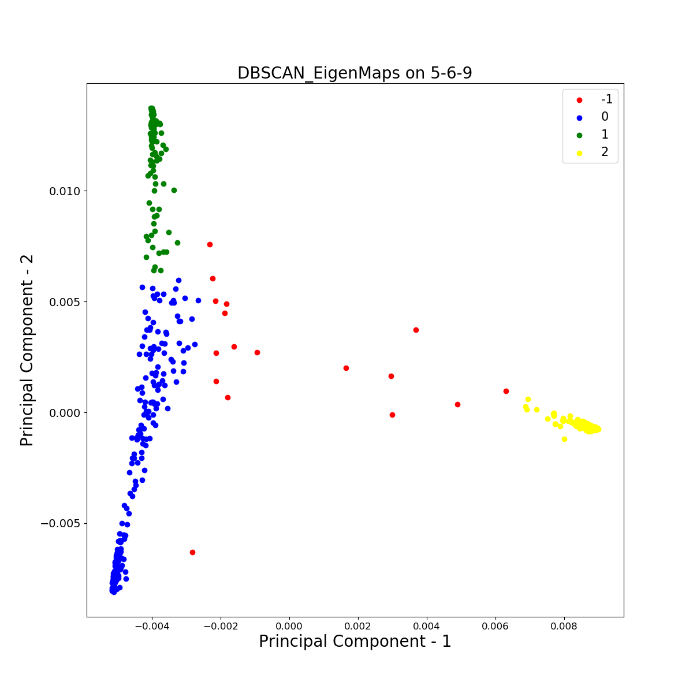
השוואה בין הייפר פרמטרים:

עבור אשכול היררכי בדקנו האם עדיף לבצע אישכול על הדאטה לפני או אחרי ביצוע רידוד המימדים

השוואה בין אלגוריתמי רידוד:

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

DBSCAN parameters:

Eps = 0.0007, min\_samples=5

DBSCAN parameters:

Eps = 10, min\_samples=22

DBSCAN parameters:

Eps = 0.01, min\_samples=10

DBSCAN parameters:

Eps = 4, min\_samples=8

DBSCAN parameters:

Eps = 1, min\_samples=19

השוואה בין הייפר פרמטרים:

עבור dbscan בדקנו:

האם עדיף לבצע אישכול על הדאטה לפני או אחרי ביצוע רידוד המימדים, ערכי eps, וmin\_samples

השוואה בין אלגוריתמי רידוד:

השוואה בין אלגוריתמי אשכול: