תרגיל מס' 3

UnSupervised learning

לספרית sklearn יש תת-ספריה של data. אנחנו ניקח משם את הדאטה של digits.

(ראה למשל: https://scikit-learn.org/stable/auto\_examples/cluster/plot\_kmeans\_digits.html)

על דאטה הזה (עשר מחלקות ו-64 תכונות) בצע הורדת מימד לשתיים (בעזרת pca או tsne). בחר 3 ספרות כרצונך – שיהוו 3 מחלקות.

הסתר את הסיווג של המחלקות.

1. הרץ אלגוריתם kmeans. נסה לפחות 2 תנאי התחלה שונים ובדוק את התוצאה של פונקצית המחיר. הרץ עבור k=1,2,…,10 ובדוק מהו ה-k הסביר בשיטת המרפק (בשאיפה לקבל 3...)
2. הרץ אלגוריתם GMM. אתחל את סיגמא (מטריצות 2\*2) למטריצה אלכסונית בה על האלכסון יש את השונות של כל תכונה (רק שתי תכונות). בדוק האם פונקצית המחיר שנלמדה בכיתה אכן יורדת אחרי כל איטרציה.

כעת השווה את תוצאותיך לסיווג האמיתי (על ידי כל מדד הגיוני שתבחר).

הגש דוח של תוצאותיך וכמובן את הקוד.

הערה: כדי לא להגיע לבעיות נומריות בהם כל האסקפוננטים שווים לאפס יש להשתמש בטריק של הלוג של הסכום. ראה למשל (במיוחד השקף האחרון שמסכם את התיקון)

<https://www.cs.bgu.ac.il/~cv201/wiki.files/logsumexp_trick.pdf>

מה שידרוש ממכם לעבוד עם ערכים מתוקנים (אך שווים) על ידי הפחתה של ערך מסוים באקספוננט.