דו"ח מסכם

תרגיל בית 3­

יונתן בתן

302279138

[Yonibettan@gmail.com](mailto:Yonibettan@gmail.com)

עמרי פרוינד

301695490

omrifro[@gamil.com](mailto:Neliahu@gamil.com)

# **חלק א'**

מבוא

# **חלק ב'**

# סעיף 1:

בקוד

# סעיף 2:

בקוד

# סעיף 3:

מדד המרחק האוקלידי מתייחס לכל אחת מהתכונות בווקטור התכונות כעל ממד במרחב. כלומר, ההשפעה של כל אחת מהתכונות הללו שוות ערך לגודלה ביחס לאחרות גם אם בפועל היחס הפוך. למשל, בעיית סיווג קשה תיווצר אם הנתונים מכילים שתי תכונות: "מעשן" / "לא מעשן" – {0,1} (לצורך הדוגמה תכונה מכרעת הקובעת כמעט לבדה את הסיווג), ו-"לחץ דם" של האדם {80-180} (לצורך הדוגמה בעל השפעה זניחה). במקרה זה מדד המרחק האוקלידי יסווג את האנשים כמעט אך ורק על פי לחץ הדם מה שבפועל יביא לסיווג שגוי. בעיה זו ממחישה את החשיבות העליונה של ביצוע עיבוד מקדים לכל dataset עליו עובדים.

האם להתייחס לנושא תכונות רציפות מול בדידות (לדעתי בעל חשיבות נמוכה בשאלה הזו)

* לדעתי אין צורך, אני השתכנעתי מהתשובה

# סעיף 4:

1. בקוד

# סעיף 5:

# סעיף 6:

# סעיף 7:

# **חלק ג'**