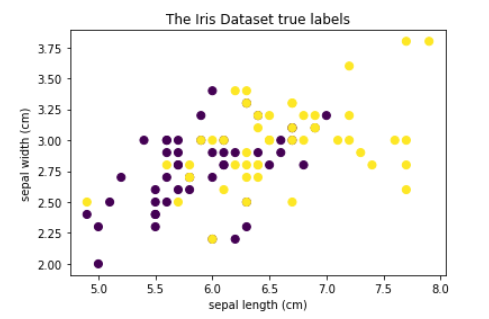
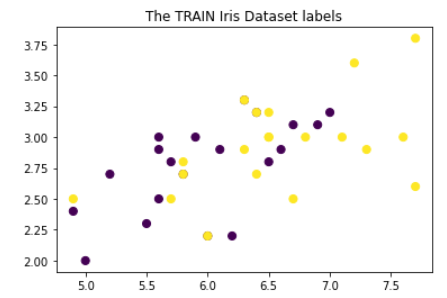
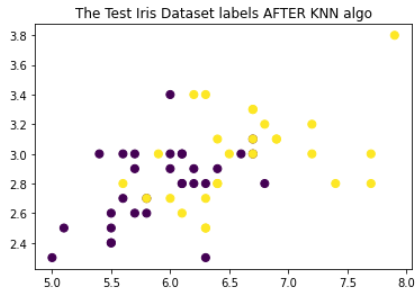
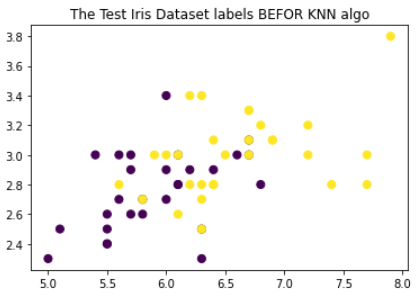
**כריית נתונים – מעבדה מס 6**

**מגישים : רותם דוידוב 318319043 | אור יהודה בן שימול 203637202**

1. Compose the Testing Dataset from the points not chosen for the Training Dataset.
2. Apply the KNN algorithm with K=1 to the Testing Dataset.
3. Show the results in scatter plots for the train and the test sets and explain the results.

Code file in the zip folder under the name lab6(6-8)

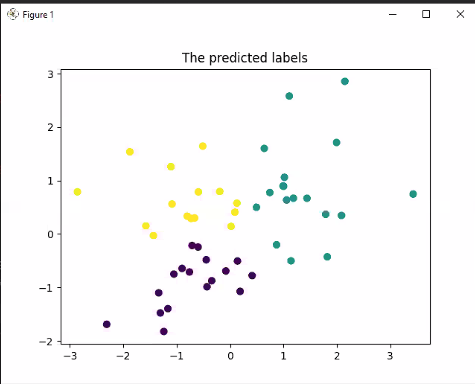
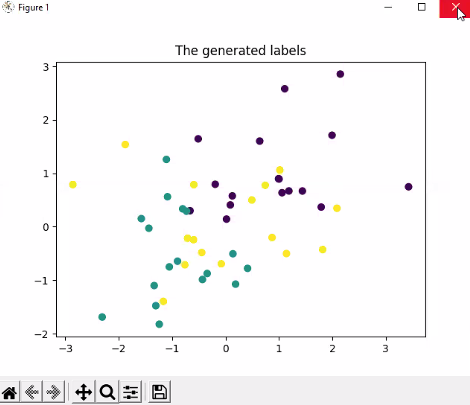


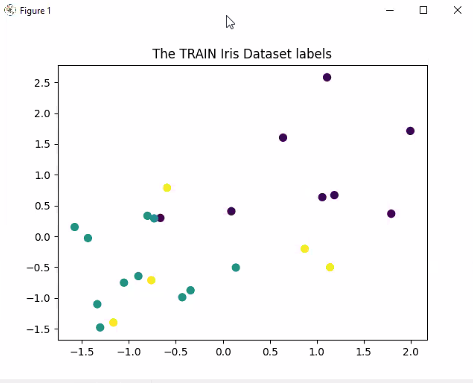
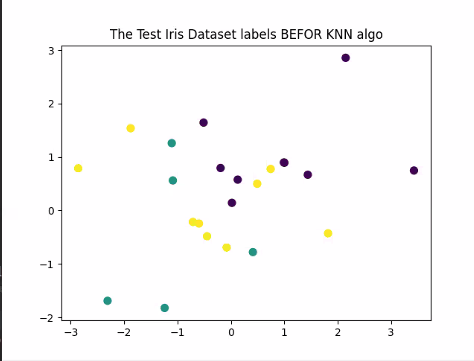


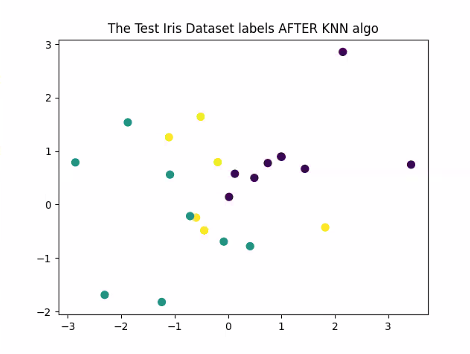
1. הגרף המכיל את הקיטלוג האמיתי של המידע
2. הגרף המכיל את קבוצת האימון
3. הגרף BEFOR מסמל איך הלייבלים של קבוצת הtest היו אמורים להראות אם KNN היה מפעיל את האלגוריתם בצורה המיטבית ביותר. (מעין extract לנתונים מהtrue רק עבור קבוצת הtest ).
4. קבוצת הtest בפועל לאחר הפעלת האלגוריתם.

על פי תוצאות הגרפים ניתן לראות כי באופן כללי האלגוריתם העניק ביצועים טובים ,

ניקח לדוגמא את הנקודה המסומנת על ידי החץ ונסביר למה שמה האלגוריתם ביצע קיטלוג דווקא לקבוצה שהנקודה לא שייכת אליה בפועל. ניתן לראות כי בפועל (גרף 3) הנקודה הייתה אמורה להסתפח לקבוצה בעלת הצבעים הצהובים, אך בפועל לאחר הפעלת האלגוריתם (גרף 4) הנקודה הסתפחה לקבוצה של הסגולים, הסיבה היא מפני שניתן לראות (בגרף2) הנקודה הקרובה ביותר לנקודה המסומנת על ידי החץ הייתה נקודה סגולה ולכן הנקודה הזאת סופחה לקבוצה הסגולה.

סעיף 9 עד 14:





אופן חיתוך הנתונים : רבעים כך ש0 עד 1/4 ו- 2/4 עד 3/4 ביחד עבור קבוצת הtrain .  
והמשלים שלהם ( 1/4 עד 2/4 ו3/4 עד סוף המערך ביחד)

באופן דומה לנאמר בסעיף קודם אפשר לראות על דוגמא של הגרף הזה ובהתאם לגרפים הבאים כי הסיפוח של הנקודה לקבוצה הסגולה ולא לקבוצה הצהובה היא משום שהשכנים הקרובים אליו ביותר היו הסגולים.

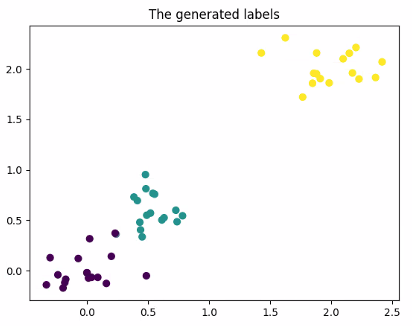
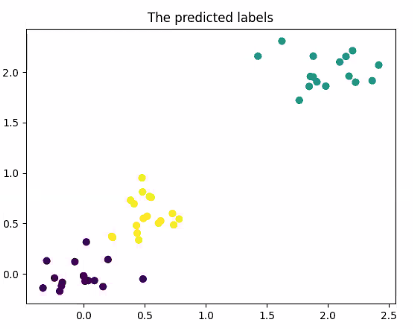
אפשר לעלות על הדעת שחלק מהקיטלוג הלא נכון נבע מחלוקת המערך וסטיית התקן שאחראית על פיזור הערכים בציר.

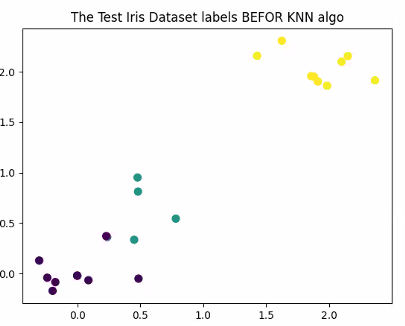
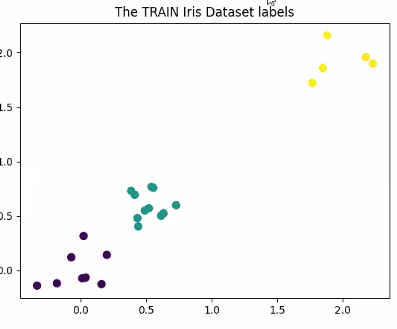


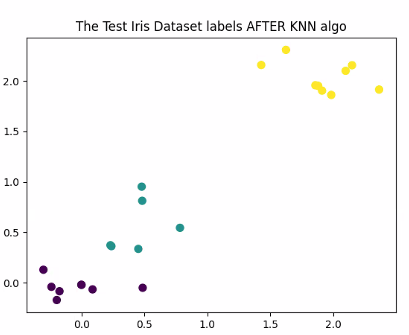
#גם פה הכותרות של הגרפים הם כמו הסעיפים 6-8

עוד 2 דוגמאות נוספות תחת אותו הסבר :



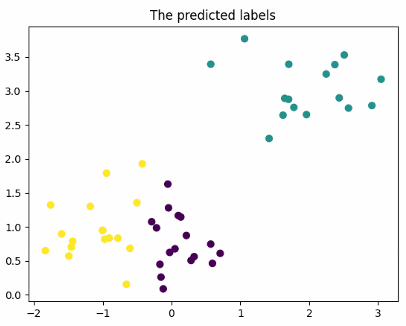
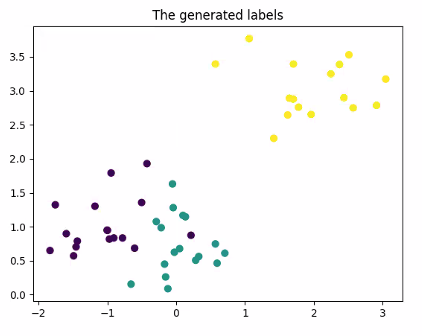
 

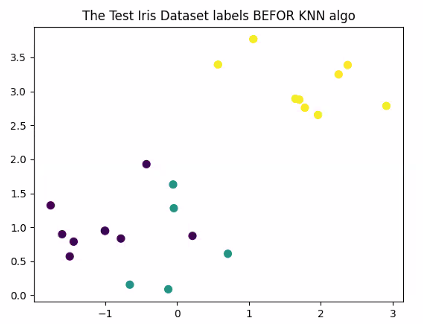
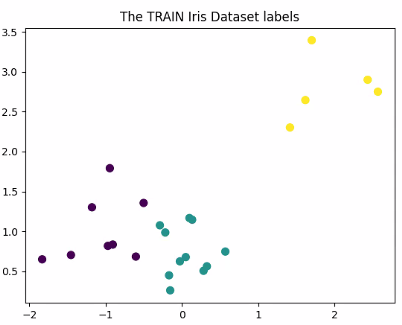


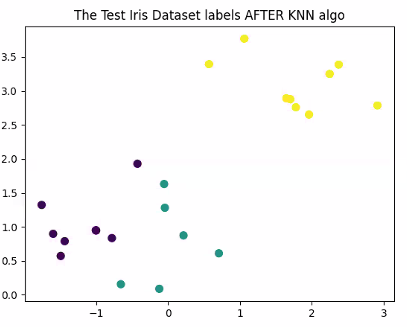


**אפשר לראות שבחלק הזה של הנתונים הקטנו את הסטיית תקן ולכן הפיזור של הנקודות מצומצם והאלגוריתם עובד בצורה מיטבית יותר מהדוגמא הקודמת.**









Code file in the zip folder under the name lab6(9-14)