**מעבדה מספר 5**

**מגישים : רותם דוידוב 318319043**

**אור יהודה בן שימול 203637202**

**סעיף מספר 1:**

Cluster the Iris Database into 2 and 3 clusters using k-means, compute the silhouette values and show the predicted labels and the silhouette results in plots (as in lab3). Which clustering is better? Explain your results.

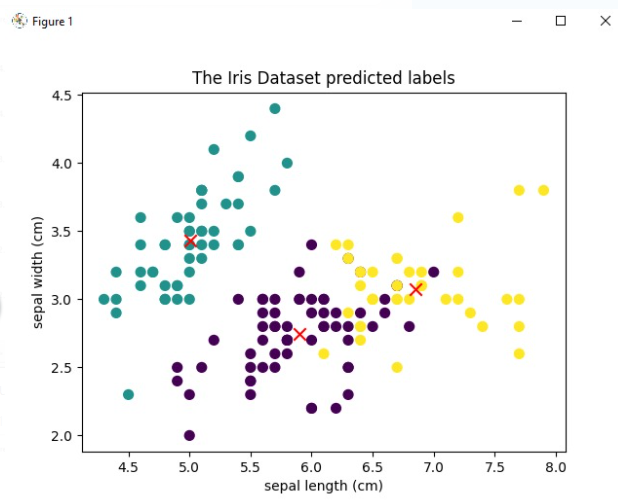
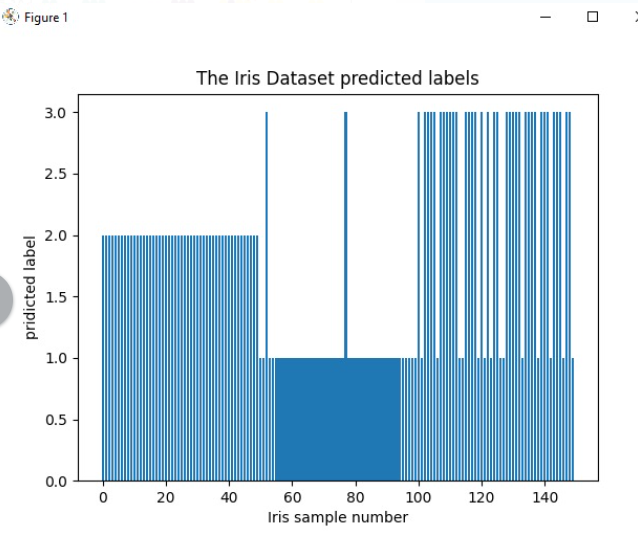
על פי התוצאות ניתן לראות כי החלוקה ל2 קלסאטרים עדיפה והתוצאות להלן :

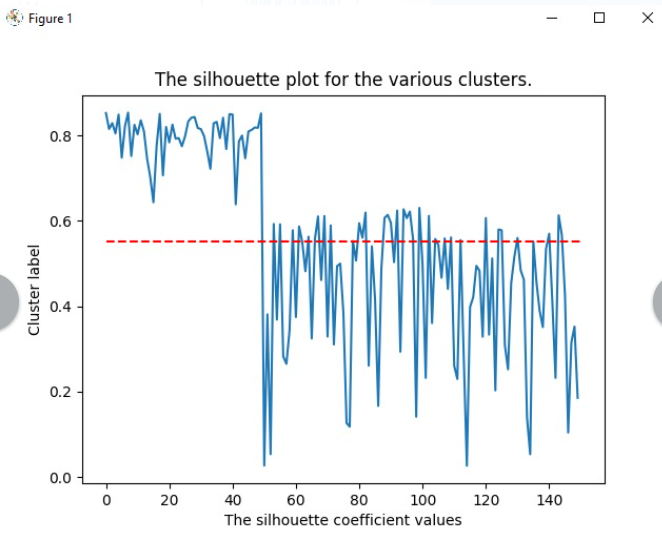
נתמקד בסילואט, ראינו במעבדות קודמות שככל שהסילואט קרוב ל1 החלוקה של הקלאסטרים טובה יותר.

ניתן לראות כי בחלוקה ל2 קלאסטרים קיבלנו סילואט בעל ערך 0.4 ולעומת זאת בחלוקה ל3 קלאסטרים קיבלנו סילואט בעל ערך 0.3 לכן מכיוון ש0.4 קרוב יותר ל1 החלוקה ל2 קלאסטרים עדיפה יותר.

עבור חלוקה ל3 קלאסטרים :

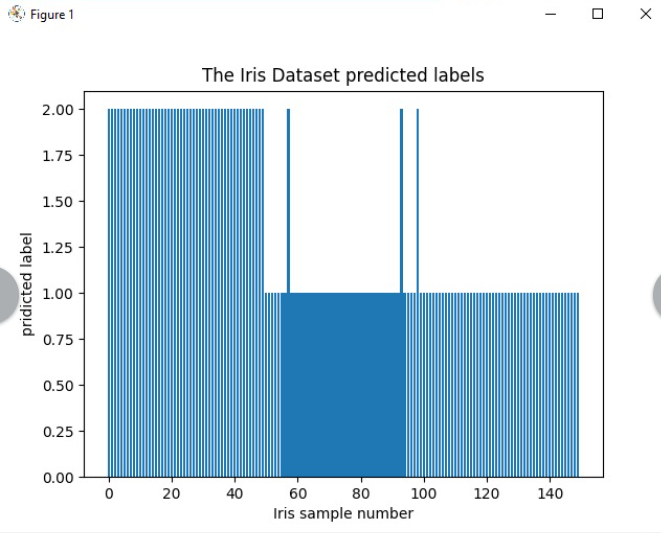
For clusters = 3 The average silhouette\_score is: 0.5528190123564095

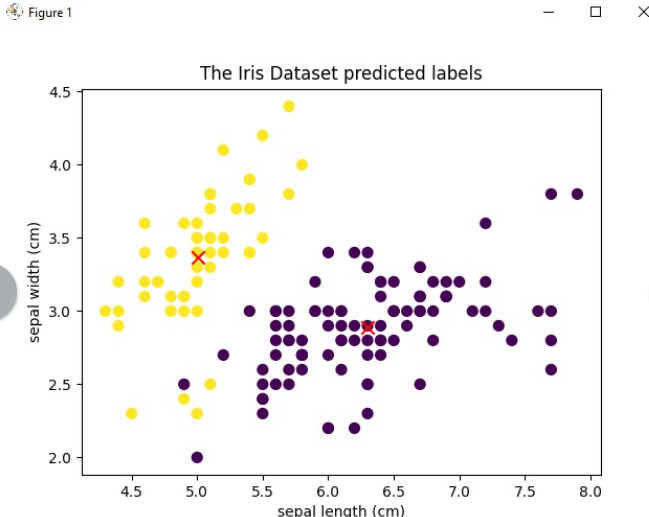
****

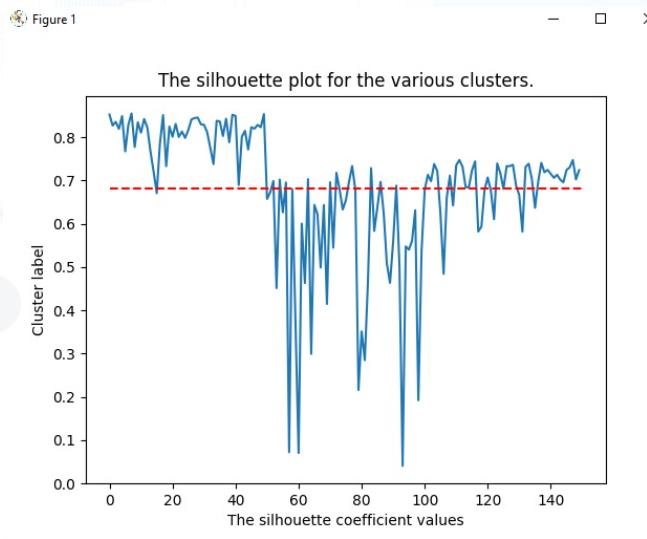
****

עבור חלוקה ל2 קלאסטרים :

For clusters = 2 The average silhouette\_score is: 0.6810461692117462

****

****

****

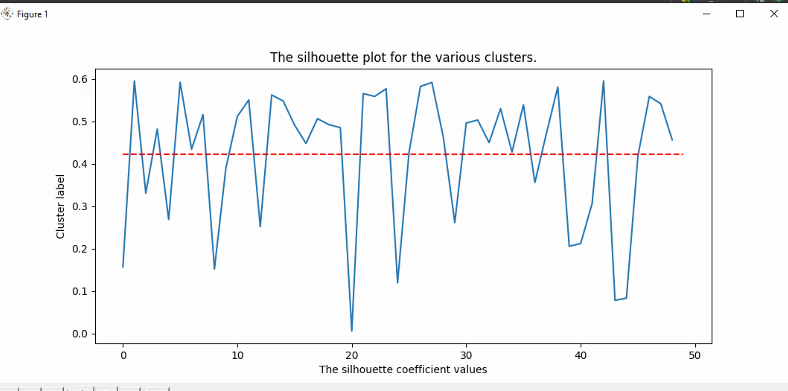
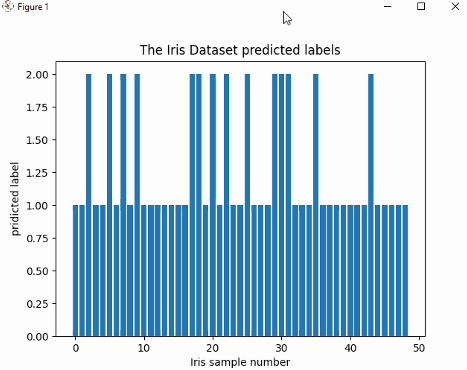
**סעיף מספר 2+3:**

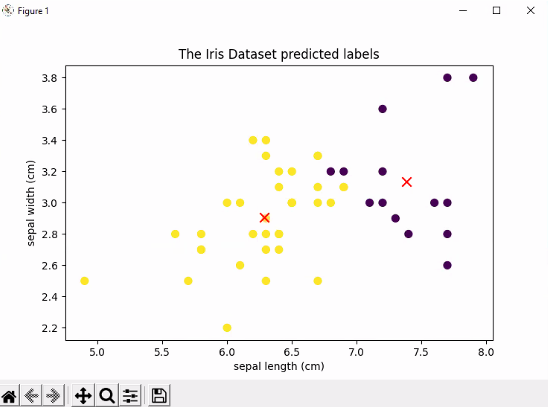
(החל מסעיף זה כל התוצאות הם על חלוקה ל2 קלאסטרים עבור טווח של 50 עד 149 מהמידע בDATASET)

Select form the IRIS data base the data of 50-149 flowers, cluster them into 2 clusters using K-means and run the prediction.

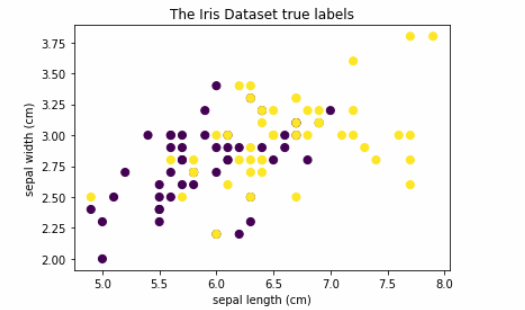
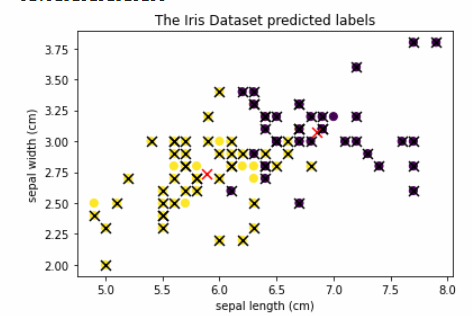
עבור חלוקה ל2 קלאסטרים :

For clusters = 2 The average silhouette\_score is: 0.44045198000771274





סעיף מספר 4

Compare the predicted results with the true labeling and mark the incorrect predicted labels in the plot with black X labels.

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

הערה(דרך לגרום ללייבלים בפרדיקט להיות קבועים ולא להשתנות) : נתנו סטייה של 60% עבור מצב שבו האלגוריתם מקלסטר בקבוצות ש1 בפרדיקטד מייצג את 2 בטרו ו2 בפרדיקטד מייצג את 3 בטרו , מכיוון שהאלגוריתם רנדומלי (הבחירה של המרכז) אז ניסינו להימנע ממצב שבו סדר הקילסטור מתהפך כלומר ש1 ייצג את 3 ו2 את 2 ולכן בלולאה הגרלנו על ידי האלגוריתם עד שהסטיה מעל 60 , באופן כללי במצב האידאלי הסטייה עומדת על 52% ובמצב הלא אידאלי(שהסיפוח הפוך ממה שאנחנו רוצים )על 98% של שגיאה.

סעיף מספר 5:

Compute the values of TP, FN, FP and TN

הערה 🡨 newColorMap=TrueLabels[50:149] 🡨

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי

ישנה אפשרות באמת לספור מהגרף עצמו אך בגלל שישנם נקודות בגרף שיושבות אחת על השנייה (למשל צהוב על סגול) אז יש נקודות שנעלמות.

סעיף מספר 6:

Calculate the values of the True Positive rate, False Positive Rate, Accuracy and Precision.

תמונה שמכילה טקסט

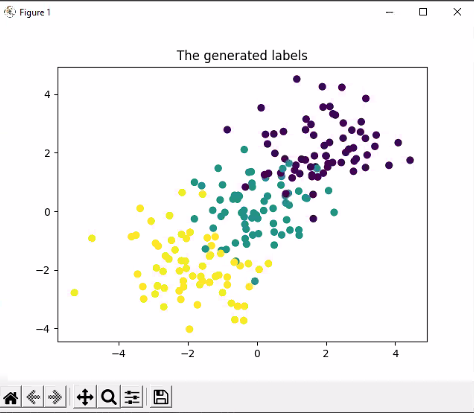
התיאור נוצר באופן אוטומטי

סעיף 8:

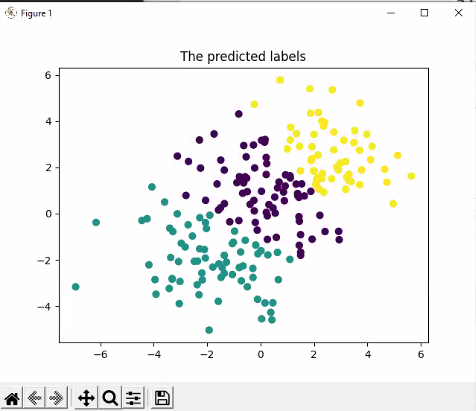
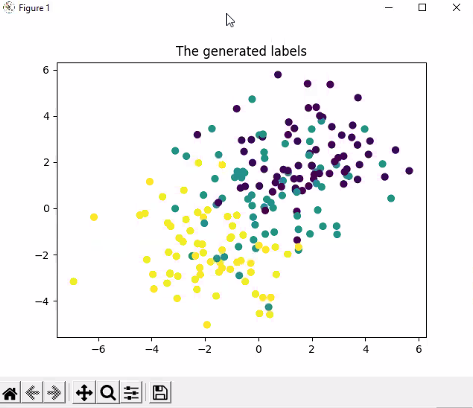


סעיף 9:

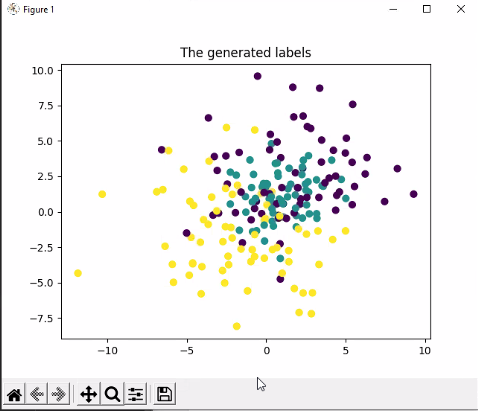
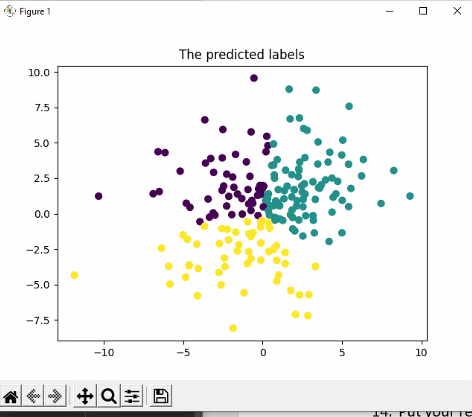
שורה ראשונה בקוד :

שורה שניה בקוד :



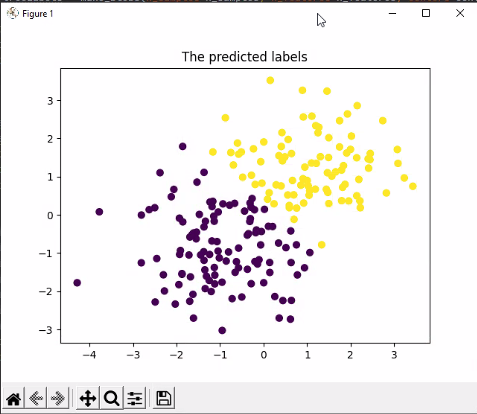
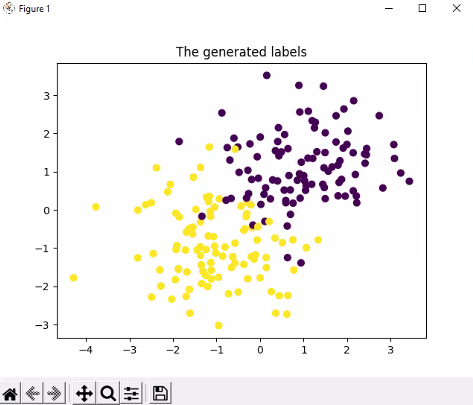
שורה שלישית בקוד :

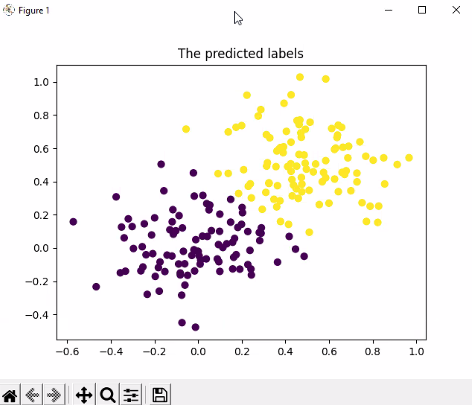
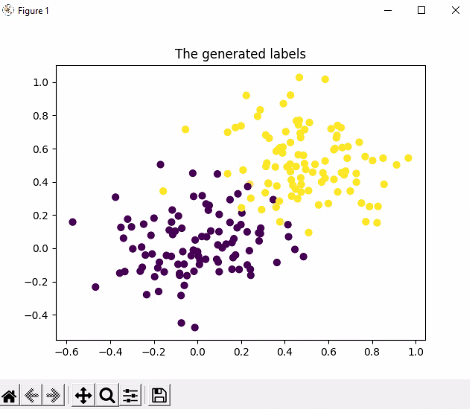
סעיף מספר 10:



שורה ראשונה של הקוד:



שורה שניה של הקוד :



סעיף 12+13

תמונה שמכילה טקסט

התיאור נוצר באופן אוטומטי