



1. נתחיל ב: dataset בשם iris. זהו dataset שמכיל נתונים שונים על פרח מזן איריס. הרעיון הוא שאתם מקבלים נתונים שונים על פרח ספציפי (פיצ'רים שלו) וצריך לנבא איזה זן של איריס זה, כלומר מה ה-class שלו.
2. הורידו את ה- iris data set - נקרא כיצד להוריד אותו מהלינק הבא: [Iris Dataset](#)
3. אנו נשתמש בפיצ'רים הראשון והשני - כלומר ניקח את שני העמודות הראשונות.
4. בנו classifier של svm כדי לסווג את הנקודות לפי הלינק הבא:  
<http://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>
5. ייצרו ארבעה "פרחים" חדשים על ידי יצירת פיצ'רים שלהם באופן מלאכותי (לבחור מספרים הגיוניים בטווח) וסווג אותם בעזרת המסווג שבניתם.
6. שרטטו בעזרת matplotlib את הנקודות כאשר כל נקודה תקבל צבע בהתאם ל class שלה, זאת על ידי קביעת הצבע של הנקודה בעזרת המשתנה c בלינק הבא: [Matplotlib scatter](#)
7. הדפיסו את ה- support vectors של ה- classifier שבניתם (רמז - בידקו מה הם השדות האפשריים של אובייקט ה- classifier). סמנו נקודות אלו באדום על אותו גרף.
8. עכשיו נשתמש בכל ארבעת הפיצ'רים האפשריים ב dataset ונשמור אותם למשתנה
9. לימדו את ה- classifier שלוש פעמים ושרטטו כל אחד מהם עם האופציות הבאות:
  - a. קרנל ליניארי
  - b. קרנל פולינומיאלי מדרגה 3
  - c. קרנל רדיאלי
10. נעשה תרגיל דומה עכשיו על ספריית הספרות: [Digits data](#)
11. הדפיסו למסך את עשר הספרות הראשונות במאגר - כדי להבין איך הספרות נראות. השתמשו בפקודה: `pyplot.imshow`
12. אמנו svm על החצי הראשון של מאגר הספרות (train data) כדי לייצר classifier של הספרות מתוך פיצ'רים שלהם. הפיצ'רים של הספרות הם כל הפיקסלים שמרכיבים אותם. כך שאם יש 64 פיקסלים לכל תמונה - אז יש 64 פיצ'רים.
13. הריצו את החצי השני של המאגר כ test, כדי לבדוק אם ה- classification יצא הגיוני וטוב ומיצאו מה ה- score שיצא.
14. שנו את ערכי C לשלושה ערכים שונים וצפו בשינוי בתוצאות
15. האם כשאנחנו משתמשים באותו ערך של C אנחנו מקבלים כל פעם אותה תוצאה? למה בריצות חוזרות התוצאה שונה? (רמז: האם יש מספרים רנדומליים בתוך החישוב?), איזה משתנה צריך לקבוע בפרמטרים של ה svm כדי שנקבל בכל פעם את אותה תוצאה?
16. מה קורה אם ה data שלנו לא מאוזן? איזה פרמטר ניתן לשנות כדי שייתן יותר משקל באימון ל classes נדירים יותר? למה?
17. איזה כיף, סיימתם לאמן את ה svm שלכם. ייתכן שהאימון על מאה אלף שורות ה data שלכם לקח עשר שעות. אתם מעוניינים עכשיו לסגור את המחשב ולהשתמש במודל שאימנתם בשבוע הבא בדמו ללקוח. כיצד תשמרו את תוצאות האימון שלכם?

זו השיטה הרשמית של הספרייה - [Model Persistence](#)

מצאנו שלפעמים בחלק מהמודלים שיטה זו לא עובדת (כנראה בגלל באגים בספרייה). במקרים כאלו צריך ממש ידנית לשמור את השדות החשובים לנו לקובץ טקסט ואז לקרוא את קובץ הטקסט ולהציב את הערכים חזרה. חישבו איזה שדות במודל (Attributes) כדאי לשמור לקובץ כדי לוודא שלא איבדנו שום מידע חשוב על ה data שלנו?

18. השתמשו ב:

<https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html> עבור iris

data set. מה היתרונות והחסרונות של פונקציה זו על פני הפונקציה הנפוצה יותר בה השתמשנו קודם? באיזה מקרים נעדיף להשתמש בזאת?

19.