

מבוא לבינה מלאכותית – תרגיל 2:

1. בשאלה זו תשתמשו ב-soft margin SVM וב-online SVM לסיווג דאטא שתייצרו בעצמכם.
- א. צרו את הדאטא לאימון, באמצעות הפונקציה `make_blobs` מתוך החבילה `scikit-learn`. כלומר, ייבאו אותה באמצעות הפקודה `"from sklearn.datasets import make_blobs"`. צרו 100 דגימות (דו ממדיות) שכל אחת מהן שייכת לאחד משני מרכזים. שימו לב שהתיוגים הנוצרים הם $\{0, 1\}$ ולא $\{\pm 1\}$.
- ב. בסעיף הזה תפתרו את בעיית הסיווג בעזרת `soft margin SVM` מתוך `scikit-learn` (היעזרו בדוקומנטציה של `scikit-learn`). ייבאו את המסווג עם הפקודה `"from sklearn.svm import LinearSVC"`. בנו את המסווג עם `hinge loss` ונסו קבועי רגולריזציה (C) שונים. הריצו את האלגוריתם למציאת הקבועים (הפונקציה `fit` של המסווג), ולאחר מכן קבלו את (\vec{w}, b) המתאימים. ציירו את הנקודות ואת הישר המסווג והסיקו מה החשיבות של C.
- ג. בסעיף הזה בנו מסווג `online SVM`. בצעו את מה שעשיתם בסעיף ב על המסווג הזה וצפו בהבדלים.

2. בשאלה זו תצטרכו לפתור את הבעיה הדואלית ל-SVM (`hard margin`) עם `kernel` שתצטרכו לבחור בעצמכם. כדי לפתור את הבעיה, השתמשו בחבילה `OSQP` הפותרת בעיות אופטימיזציה ריבועיות, או בכל `quadratic programming solver` אחר שתמצאו. נתון (ב-math wiki) קובץ אקסל שבו שני גיליונות. בכל אחד מהם יש דגימות מתויגות, המופיעים בצורה הבאה:

עמודה 1	עמודה 2	עמודה 3
קואורדינטה x	קואורדינטה y	תיוג 0/1

לכל סט דגימות, מצאו `kernel` מתאים ופתרו את הבעיה. ציירו את המסווג (שהפעם לא יהיה לינארי) ואת הדגימות במרחב המקורי.