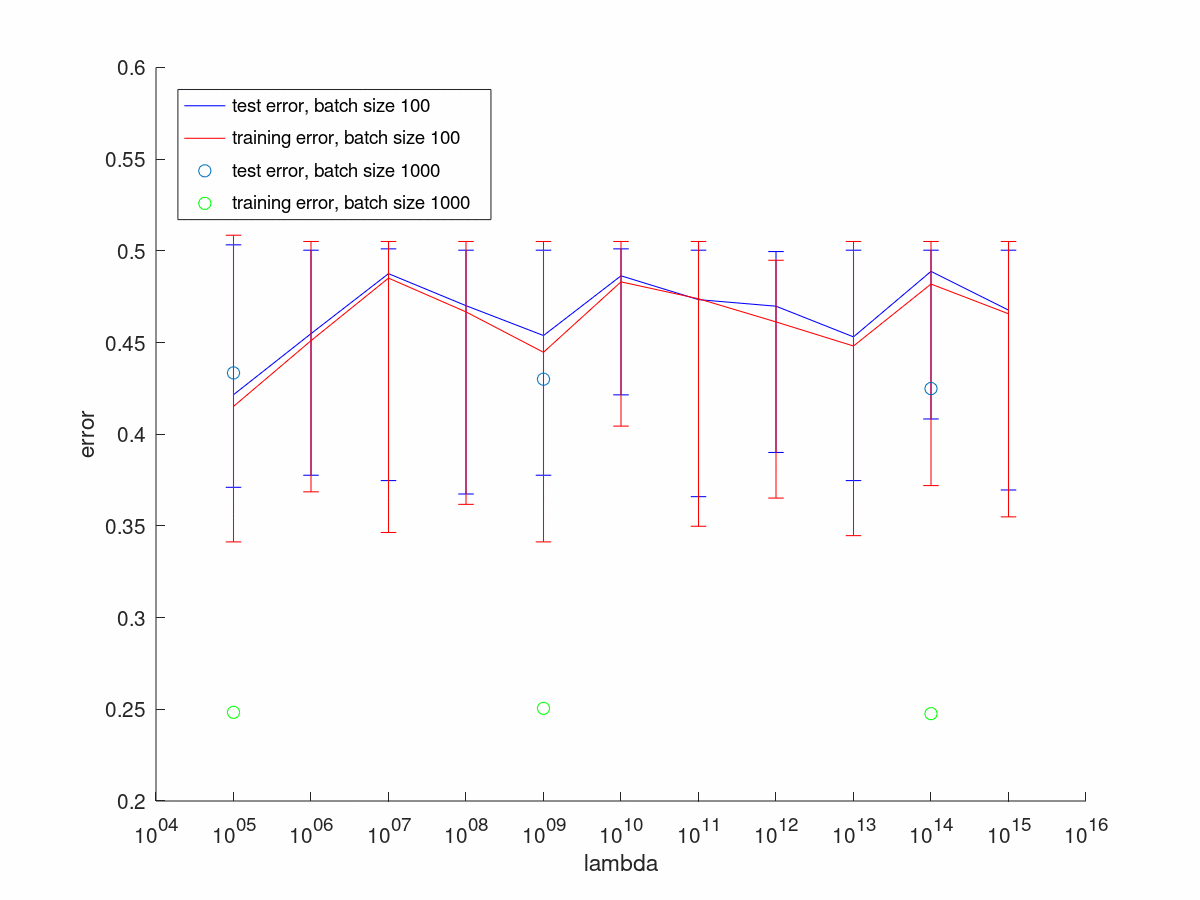
עבודה 2 – מבוא ללמידה

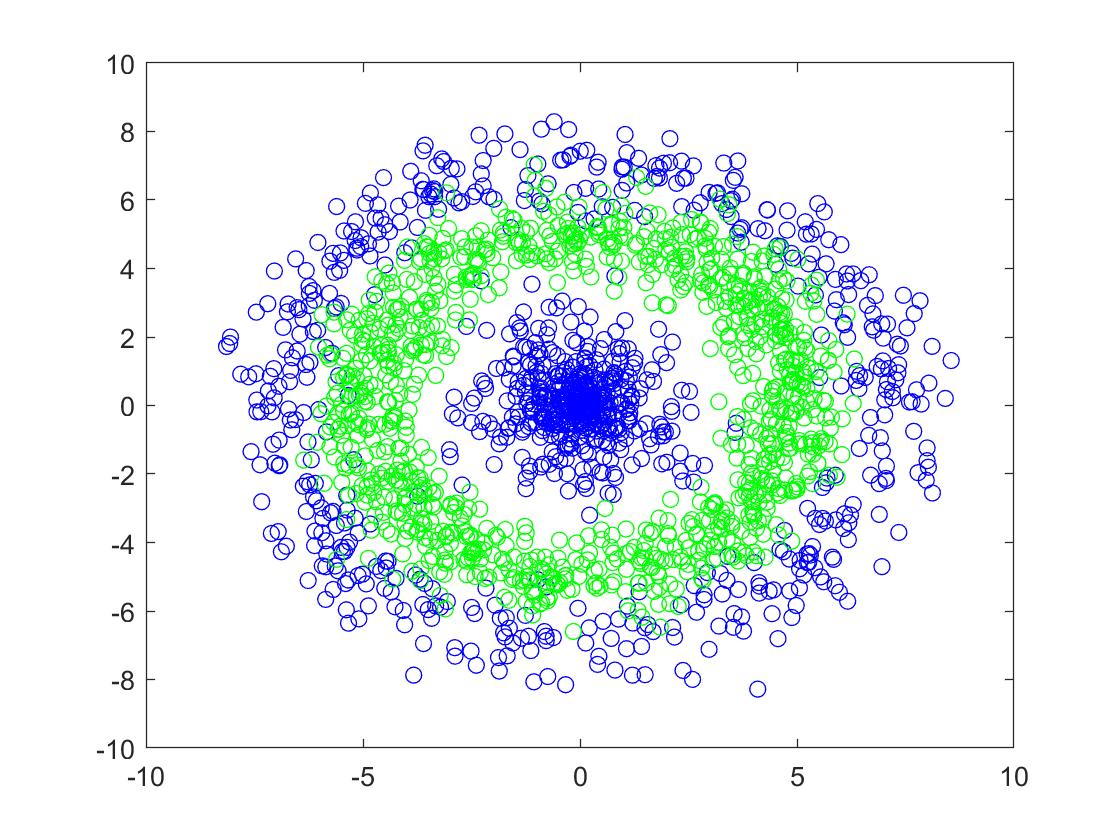
**מגישים: רעי וייס-ליפשיץ ועומרי אטל**

שאלה 2



1. אנו מצפים שהניסוי הראשון (בעל המדגם הגדול יותר) יהיה בעל training error וגם test error נמוכים יותר מאשר הניסוי השני (הניסוי בעלי המדגם הקטן). התוצאות משקפות את ההפרש בין ה-training error, אך לעומת זאת ב- test error אנו מקבלים הפרש שגיאות נמוך יותר בין הניסויים. יתרה מכך, בניסוי הראשון אנו מקבלים שגיאה גבוהה יותר בלמבדה הנמוכה, בניגוד לציפיות.

**שאלה 4**

1. 

ניתן לראות שנקודות המדגם לא מותאמות להפרדה לינארית על ידי soft SVM. נרצה להעלות את מימד המדגם ובכך לקבל מפריד לינארי טוב יותר, ונשתמש ב – kernel soft SVM בשביל להקטין את סיבוכיות החישוב.

שאלה 5

נניח בשלילה שקיימת כך ש- . משום ש- , אז עבור נקבל ש- . אז עבור נקבל ש -  
 בסתירה לחוקי מכפלה פנימית ולכן אין כזו.

1. נניח בשלילה שקיימת כך ש- . נגדיר ו-  
    . לכן מתקיים ש- בסתירה לחוקי מכפלה פנימית ולכן אין כזו.
2. נתבונן בבעיית האופטימיזציה

נראה שבעיה זו שקולה לבעיה מהסוג:

ואז לפי משפט הייצוג ינבע שקיים פתרון w לבעיה המקורית שניתן לייצג כקומבינציה לינארית של -ים עבור שהיא פונקציית הזהות, ומכאן w יהיה קומבינציה לינארית של כל ה -ים.

תחילה נגדיר . וברור ש- מונוטונית לא יורדת.

נגדיר . אז נשים לב שעבור שהיא פונקציית הזהות נקבל ש-

1. *נראה שקיימת כך ש-*

*נגדיר כך ש- . אזי:*

שאלה 6

1. עבור , נקבל מחלקת היפותזות קטנה יותר מ- , ולכן שגיאת ה-approximation שלנו יכולה לגדול. במילים אחרות, יכול להיות שאין לנו כלל מספיק טוב במחלקת ההיפותזות הקטנה יותר. לעומת זאת, עבור , נקבל מחלקת היפותזות גדולה יותר מ- . לכן, במקרה זה, שגיאת ה-approximation תקטן וכתוצאה מכך יכול להיות שנקבל שגיאת estimation גדולה יותר. כלומר יכול להיות שיש לנו כלל שמבצע overfitting על המדגם.
2. נסביר תחילה ש- הינה realizable לפי , שכן לפי נתוני השאלה, הכלל שבו הרופאים משתמשים כולל בתוכו 10 תווים לכל היותר, ומשום ש- מייצג את החלטות הרופאים אז קיימת היפותזה ב- שהכלל שלה הוא הכלל של הרופאים ולכן בעלת *שגיאה 0 על ההתפלגות .*

*נמצא חסם עליון K, כך שלכל מדגם S רנדומלי בגודל יתקיים שאלגוריתם ה ERM יחזיר לנו היפותזה בעלת שגיאה בהסתברות לפחות 0.99.*

*לפי משפט PAC שנלמד עבור שהיא* realizable *לפי ועבור מתקיים שלכל מדגם רנדומלי S בגודל m כך ש- , כל אלגוריתם ERM יחזיר לנו היפותזה שמקיימת בהסתברות של לפחות .*

*נמצא חסם עליון ל- .*

*ולכן בפרט, לכל מדגם בגודל לפחות*

*מתקיים הדרוש לפי משפט PAC.*

1. *כעת, לא נתון ש- היא* realizable *לפי עבור , שכן ייתכן שאין כלל שקול לכלל הרופאים באורך n. נמצא חסם עליון K, כך שלכל מדגם S רנדומלי בגודל יתקיים שאלגוריתם ה ERM יחזיר לנו היפותזה בעלת שגיאה עודפת 0.1 בהסתברות לפחות 0.99.*

*לפי משפט PAC עבור המקרה האגנוסטי עבור ו- , ו- , לכל שלכל מדגם רנדומלי S בגודל m כך ש- , כל אלגוריתם ERM יחזיר לנו היפותזה שמקיימת בהסתברות של לפחות  
 .*

*נמצא חסם עליון ל- . בדומה לסעיף א' ולכן בפרט, לכל מדגם בגודל לפחות*

*מתקיים הדרוש לפי משפט PAC.*

שאלה 7

נמצא תוכנית ריבועית המתאימה לבעיית ה- soft-svm המוצגת בשאלה, נסמנה (שהיא בעיית המינימיזציה של )

כאשר .

נגדיר בעיית מינימיזציה חדשה, נסמנה , ונראה שפתרון שלה הוא פתרון למינימיזציה .

כך ש-

תחילה נוכיח שפתרון של , הינו פתרון ל-.

נסיק מהאילוצים של ש

hence the optimal value for

לכן הבעיה שקולה לבעיה , ולכן שפותר את יפתור את .

כעת נמצא ייצוג של בעיית כתוכנית ריבועית. נמצא שיהיו פרמטרים לתוכנית הריבועית.

, s.t

, s.t the amount of 1’s is m and the number of zeros is 3m.

שאלה 8

אבחנה 1: לכל מתקיים ש- . שכן וגם  
 וגם .

אבחנה 2: לפי הגדרת המדגם S מתקיים לכל ש- .

1. יהי

נוכיח שלכל איטרציה של אלגוריתם ה Perceptron עבור המדגם S, . נוכיח זאת באינדוקציה על t.

מקרה בסיס: . אז לפי אתחול האלגוריתם .

נניח את הטענה עבור t-1, כלומר נניח ש- ונוכיח עבור t.

צעד: יהי t מס' טבעי.

*כאשר לפי הנחת האינדוקציה.*

1. *יהי המפריד שמוצא אלגוריתם ה* Perceptron. *נוכיח שלכל i מתקיים באינדוקציה על i.*

*מקרה בסיס: . משום ש- הינו המפריד שחוזר מהאלגוריתם אז לפי משפט שנלמד בכיתה מתקיים שלכל . בפרט הטענה נכונה עבור דוגמה  
 ותוויתה . אז נשים לב שבהכרח . בנוסף, לפי אבחנה 1 ומכלל עדכון ה-*Perceptron *נובע ש- .*

*נניח את הטענה עבור לכל , כלומר נניח ש- .*

*צעד: יהי . נראה שמתקיים . נתבונן בדוגמה ובפרט ב-*

*\* לפי אבחנה 2.*

*נעביר אגפים ונקבל לפי הנחת האינדוקציה ש-*

1. *נראה שמספר האיטרציות של אלגוריתם ה* Perceptron *עבור מדגם זה הינו אקספוננציאלי במימד d. לפי סעיף א עבור מתקיים ש- , כלומר . בנוסף, לפי סעיף ב מתקיים ש- ואז סה"כ נקבל ש-  
   . כלומר זמן הריצה אקספוננציאלי ב-d.*

*שאלה 9*

*נסמן . נגדיר   
ו-. אזי וגם לכל מחוקי נגזרת מתקיים ש- .*

1. *לפי הגדרת אלגוריתם Gradient Descent מתקיים ש- . יהי נחשב את על ידי מציאת .*

*ולכן מתקיים ש-*

*סה"כ נקבל ש-*

1. *לפי הגדרת אלגוריתם Stochastic Gradient Descent עבור רנדומלי*

*לפי חישובים שביצענו בסעיף א:*

*ולכן סה"כ*