

למידה חישובית - תרגיל 2 - משימה מעשית

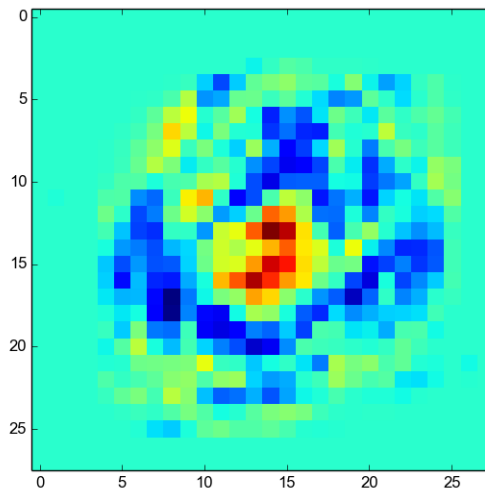
חגי ב.י., יונתן צ.

1 -

(א) טבלת דיוק לפי מספר דגימות n :

n	accuracy	percentile %5	percentile %95
5	0.910250768	0.833162743	0.933469806
10	0.899631525	0.831806551	0.940122825
50	0.953075742	0.919447288	0.975486182
100	0.969646878	0.951842375	0.984672467
500	0.978561924	0.955450358	0.990276356
1000	0.985808598	0.978352098	0.990839304
5000	0.988137155	0.976432958	0.99283521

(ב) תמונת w - המשקולות של הפרסטרון



ניתן לראות שבמרכז (איפה שאדום) ישנם ערכים גבוהים. כלומר העל-מישור המתאים ל- w נוטה לסווג תמונות שבמרכזן יש פיקסלים לבנים (ערכים גבוהים) כתמונות מלייבל +1, כלומר הספרה 8 (ולא הספרה 0). זה הגיוני משום שבמרכז הספרה 8 ישנה הצטלבות (כלומר פיקסלים לבנים במקרה של התמונות מ-MNIST original שהן לבן על גבי שחור), ובספרה 0 ישנו חלל ריק (כלומר פיקסלים שחורים). ברוב התמונות ה-8 במדגם כנראה שההצטלבות אכן נמצאת קרוב למרכז התמונה.

(ג) הדיוק של הפרסטרון על ה test set (לאחר שאומן על כל ה-train set) הוא: 0.991299897646



(ד) תמונה שסווגה לא נכון (כ-8):

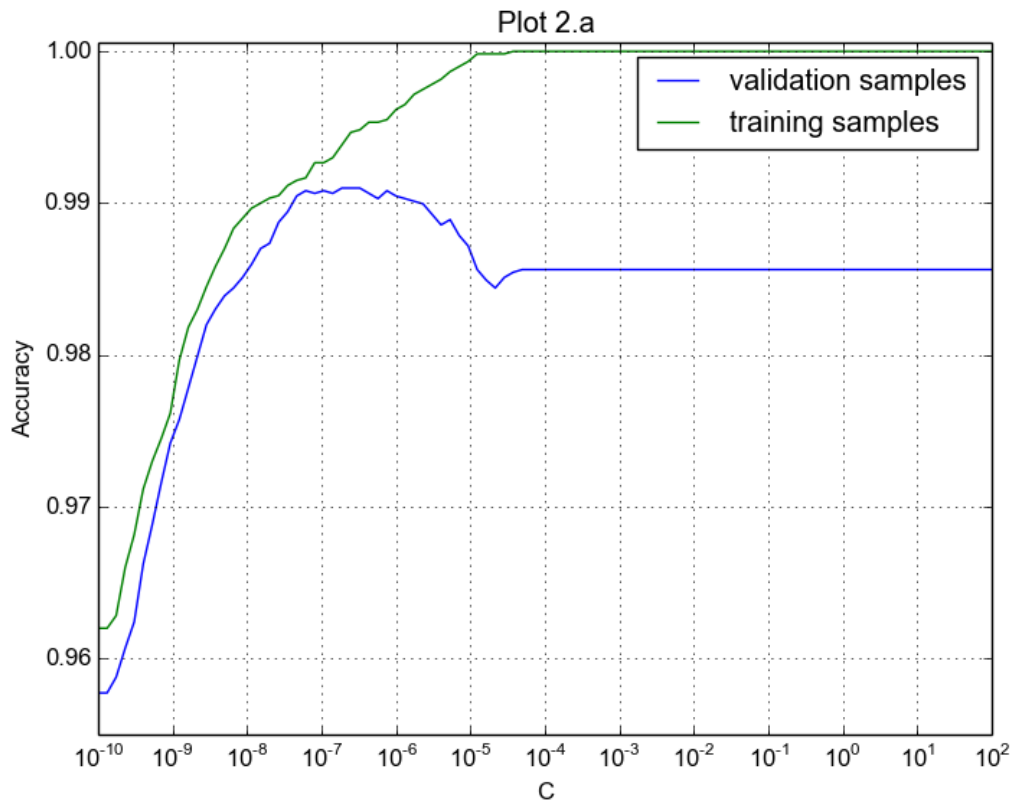
ניתן להסביר את הסיווג השגוי ע"י כך שבמרכז התמונה ישנם פיקסלים לבנים, וכאמור בסעיף ב' זה גורם לפרסטרון לסווג אותה כ-8 (בטעות במקרה הזה).



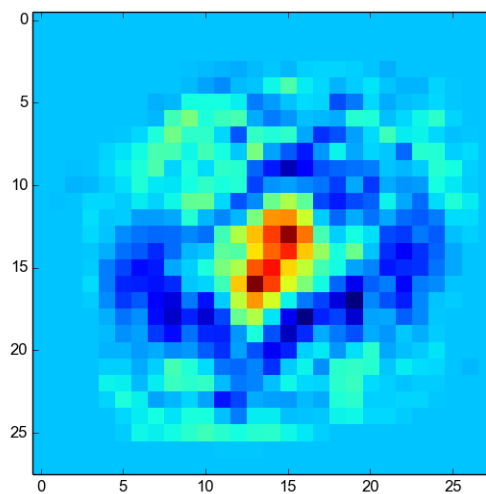
תמונה שסווגה לא נכון (כ-0):

ניתן להסביר את הטעות ע"י כך שהצטלבות של ה-8 נמצאת גבוהה (יחסית לדגימות 8 אחרות), ולכן הפיקסלים במרכז התמונה שחורים, ולכן הפרסטרון מסווגה באופן שגוי כ-0.

2. (א) גרף דיוק (על מדגם אימון, וידוא) כפונקציה של פרמטר C :

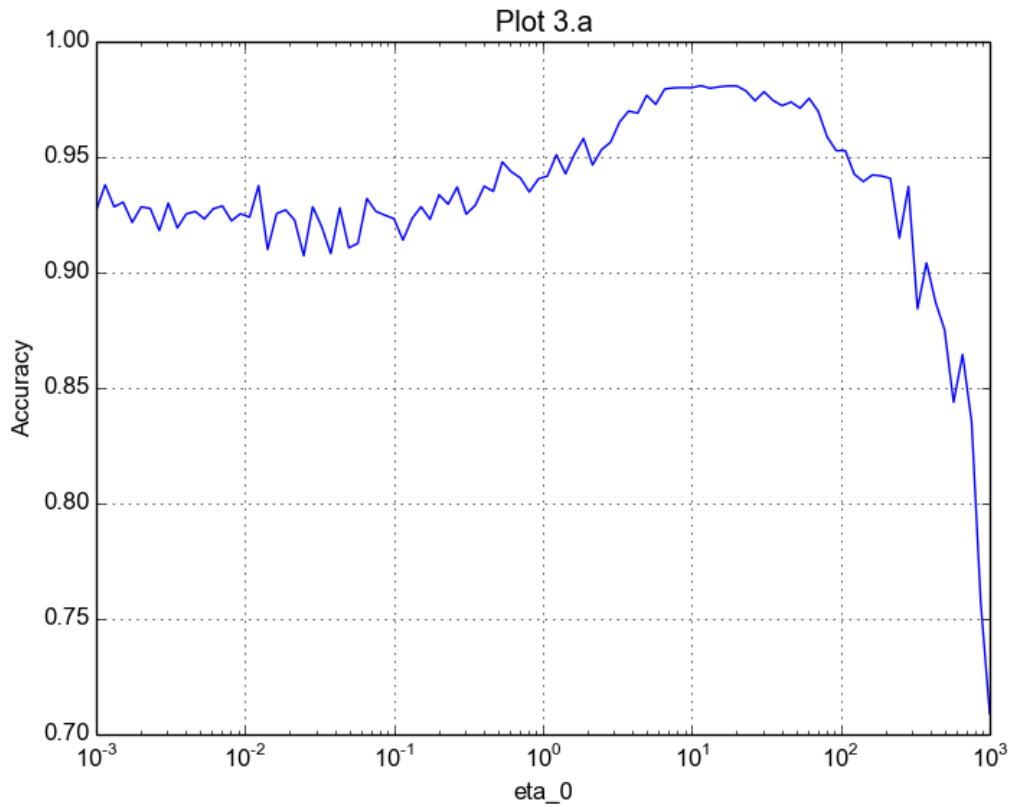


ה- C שבוא התקבל דיוק מקסימלי הוא $C = 1.87381742286 \times 10^{-7}$, והדיוק שהתקבל הוא: 0.990994111534
 (ב) הפרמטר C קובע כמה אנו "מענישים" שגיאות בסיווג של דגימות האימון (ככל ש- C גדול יותר המסווג "יעדיף" פחות שגיאות סיווג, בעלות של margin קטן יותר). לכן אנחנו רואים שככל שמגדילים את C ה- C training accuracy גדל, כלומר השגיאה על ה training set קטנה (כי כאמור המסווג יבצע פחות טעויות סיווג), עד אשר בנקודה מסוימת היא מגיעה לשגיאה 0 (בערך ב- $C = 10^{-5}$).
 מנגד, בגלל שערכי C גדולים מדי מובילים למסווג עם margin קטן, אנו רואים שהחל מ- $C = 1.87 \times 10^{-7}$ יכולת ההכללה של המסווג נפגעת, כלומר ה- C validation accuracy יורד.
 (ג) תמונת w - המשקולות של המפריד:

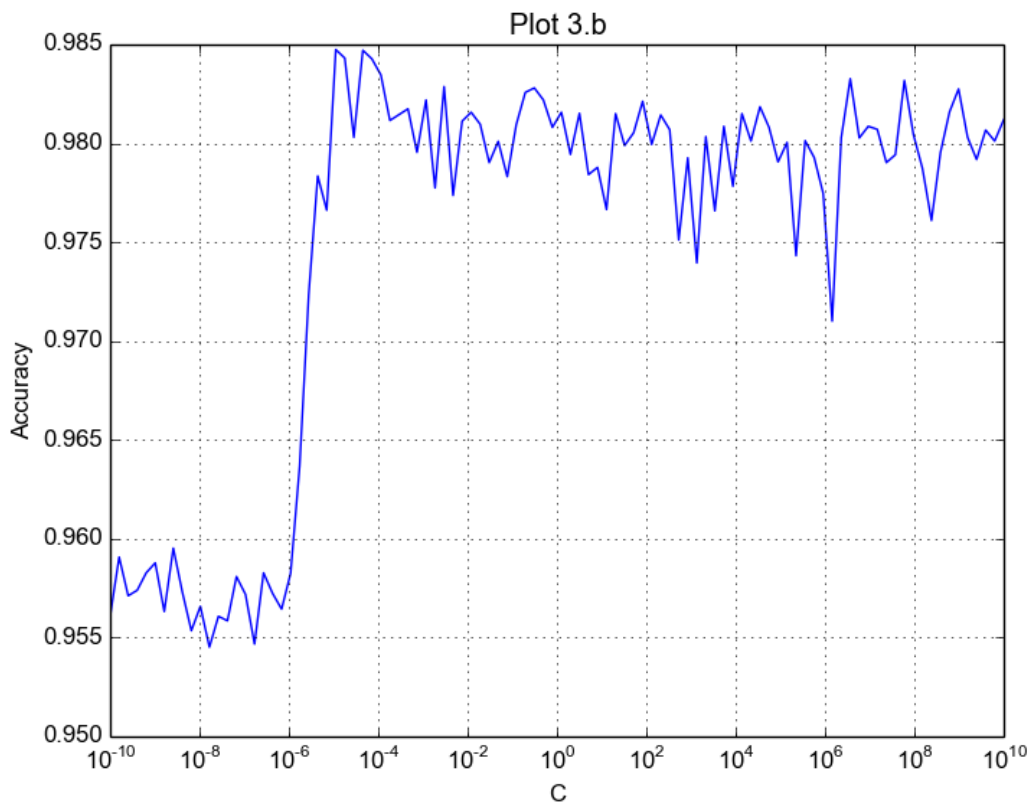


(ד) הדיוק של המסווג הכי טוב על ה- C test set הוא: 0.992323439099

3. (א) גרף דיוק (ממוצע) כפונקציה של הפרמטר η_0 :

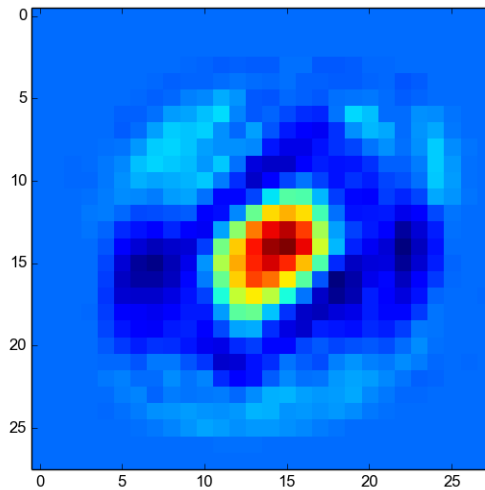


הדיוק המקסימלי התקבל ב- $\eta_0 = 11.497569954$, והוא: 0.980897125043.
 (ב) גרף דיוק (ממוצע) כפונקציה של הפרמטר C :



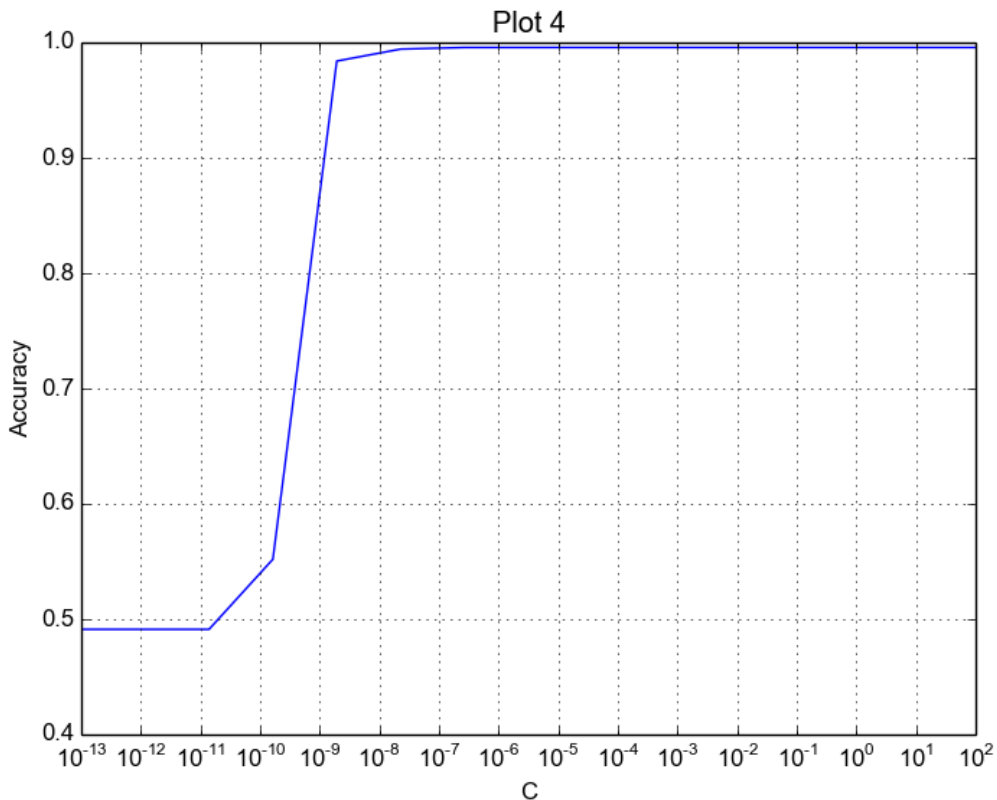
הדיוק המקסימלי התקבל ב- $C = 1.12332403298 \times 10^{-5}$, והוא: 0.984741946657.

(ג) תמונת w - המשקולות של המפריד:



(ד) הדיוק של המסווג הכי טוב על ה-`test set` הוא: 0.991811668373

4. אכן הצלחנו למצוא מסווג SVM שיתן תוצאות יותר מדויקות מהתוצאות הקודמות. השתמשנו ב-SVC (לא LinearSVC) שב-`sklearn` בצורה דומה לשאלה 2, אבל עם שימוש ב-`quadratic kernel`, כלומר ב- $K(x, x') = (1 + x \cdot x')^2$. גרף דיוק של SVC עם קרנל כפונקציה של C :



התוצאה הטובה ביותר התקבלה ב- $C = 2.68269579528 \times 10^{-7}$, והיא: 0.99567024593. הדיוק של המסווג הכי טוב (עם ה- C הנ"ל) על ה-`test set` הוא: 0.996417604913, שזה משמעותית יותר טוב מהמסווגים האחרים. אנו מניחים ששימוש בקרנל גאוסיאני יכול להביא לתוצאות אפילו יותר טובות, אבל ההרצה שלו איטית, וצריך לעשות חיפוש Grid על ערכי C, σ^2 , לכן בחרנו בקרנל הפשוט ביותר שיכולנו.