

פס: טלוח יבוקין

בתרגיל זה תתנסו במימוש האלגוריתם BackProp ושימוש באלגוריתם זה לזיהוי מספרים הכתובים בכתב יד.

כחלק מהתרגיל מצורף קובץ Matlab המכיל 5000 דוגמאות של מספרים הכתובים בכתב יד. הדוגמאות נלקחו ממאגר MNIST

<http://yann.lecun.com/exdb/mnist>

הקובץ מורכב משני משתנים האחד X - משתנה 5000 עמודות כשכל עמודה היא מספר בכתב יד המצויר על גבי תמונות של 20 על 20 פיקסלים.

המשתנה השני Y מכיל את המספרים המצוירים בעמודות המתאימות במשתנה הראשון.

א. ציירו את המספר הנמצא בעמודה שרירותית ובדקו שאכן זה מתאים למספר המתאים ב Y. ע"מ לצייר את המספרים בשחור ולבן אנא נרמלו לגבי כל מספר (עמודה ב X) את הערכים לפי הערך המקסימלי בעמודה זו. ב - Matlab ניתן להשתמש ב colormap(gray) עבור הצבעים ו ב imagesc. ו [-1,1] היא הסקלה של השחור ולבן לאחר הנרמול. תמונת הציור שהתקבל יש להגיש.

ב. לרשת הניורונים שאתם אמורים לבנות יש 400 כניסות (כמספר הפיקסלים) + Bias, בשכבת הכניסה. 25 נייורונים בשכבה האמצעית + Bias. בשכבה האחרונה יש 10 נייורונים שאלו 10 המספרים האפשריים. בנו את הרשת והשתמשו באלגוריתם BackProp ע"מ לבנות את הנגזרות המשמשות למציאת המינימום של פונקציית המחיר/פוטנציאל. זכרו שהפונקציה שלכם אמורה לקבל את הפרמטרים של הרשת כווקטור אחד ארוך להעביר אותו לקונפיגורציית הרשת כלומר לפי הטופולוגיה של הרשת. תוצאת הפונקציה צריכה להיות שני משתנים. אחד זה ערך הפוטנציאל עבור הפרמטרים המתונים והשני זה ווקטור הנגזרות (ווקטור אחד ארוך) של הפוטנציאל לפי הנגזרות. כלומר בסוף חישוב ה BackProp יש להמיר את הנגזרות מטופולוגיית הרשת חזרה ל ווקטור ארוך. השתמשו ב - reshape.

ב1. בצעו הרצה אחת של ה-backProp והשוו את התוצאה לחישוב נומרי של נגזרות המבוצעת באופן ישיר. וודאו שהתוצאות קרובות וציירו את ההפרשים

את התוצאות. כמובן שיש לאתחל את הרשת בצורה אקראית לפני החישוב הראשון.

ב2. תוך שימוש בפונקציית ה-BackProp שבניתם ופונקציה מוכנה אמנו את רשת הניורונים כשאתם משתמשים ב 60% מה Data העומד לשרותכם - Trainig set.

כיוון שלבעיה זו הרבה מאפיינים הפונקציה שהומלצה בכיתה fminunc יכולה לקרוס ולכן עדיף להשתמש בפונקציה fmincg. לפונקציה זו צריך להעביר את הקריא ל BackProp שלנו שמחשבת נגזרות ואת הפוטנציאל עצמו, פרמטרים התחלתיים של הרשת (בצורת ווקטור) ו options שכל מה שחשוב עבורם זה פרמטר ה MaxIter. ניתן לשנות אותו ישירות בפונקציה עצמה. ציירו את ערך פונקציית המחיר/פוטנציאל כפונקציה של פרמטר זה כשהוא גדל מ - 10 עד ל 500.

ג. שנו את חישוב ה BackProp כך שיקלול את הרגולריזציה. כלומר הוסיפו רכיב הרמוני לפוטנציאל שלא מאפשר לפרמטרים לגדול בצורה לא מבוקרת. את השינוי הזה ניתן לעשות בסוף הפונקציה כשכבר הפכתם מקונפיגורציית רשת לווקטור. זיכרו לחלקי ה bias לא מבצעים רגולריזציה.

ג1. עבור ערכים הולכים וגדלים של פרמטר הרגולריזציה (למבדה) אמנו את הרשת הניורונים (MaxIter גבוה) ובדקו מה הוא הערך של פונקציית מחיר/פוטנציאל כשמחשבים אותו תוך שימוש ברשת המאומנת המופעלת על CrossValidation data כ - 20% מה DATA.

ג2. ציירו גרף של הערכים הניל כפונקציה של למבדה. האם קיים ערך אופטימלי? 2. תוך שימוש בערך האופטימלי חשבו מה הערך של הפוטנציאל עבור 20% הנתונים הנותרים ומה אחוז ההצלחה של הרשת.

ד. ללא חישובים מפורטים הציעו ותארו מנגנון לזיהוי מספרים בכתב יד אשר משתמש ב Vector Supporting Machines.

הורדת תוכן הקבצים

הורדת תוכן הקבצים

- trgl.mat

figure 1

hand-written digits:

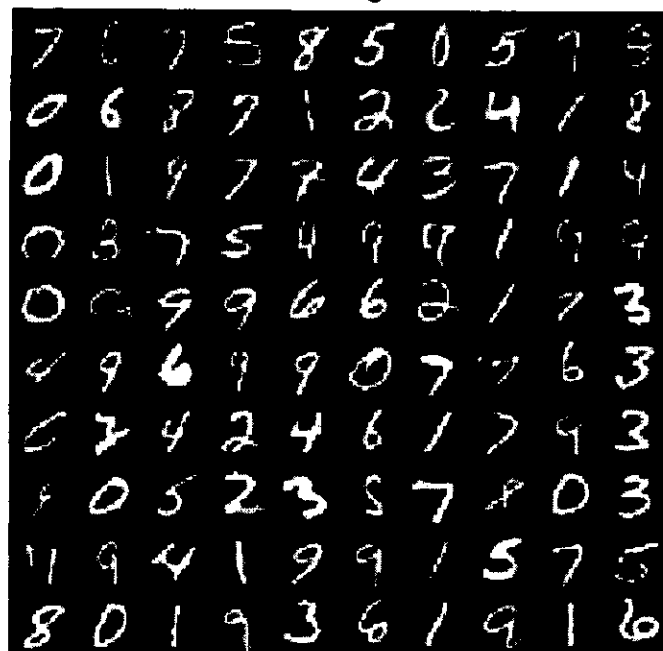


figure 2

Recognized numbers:

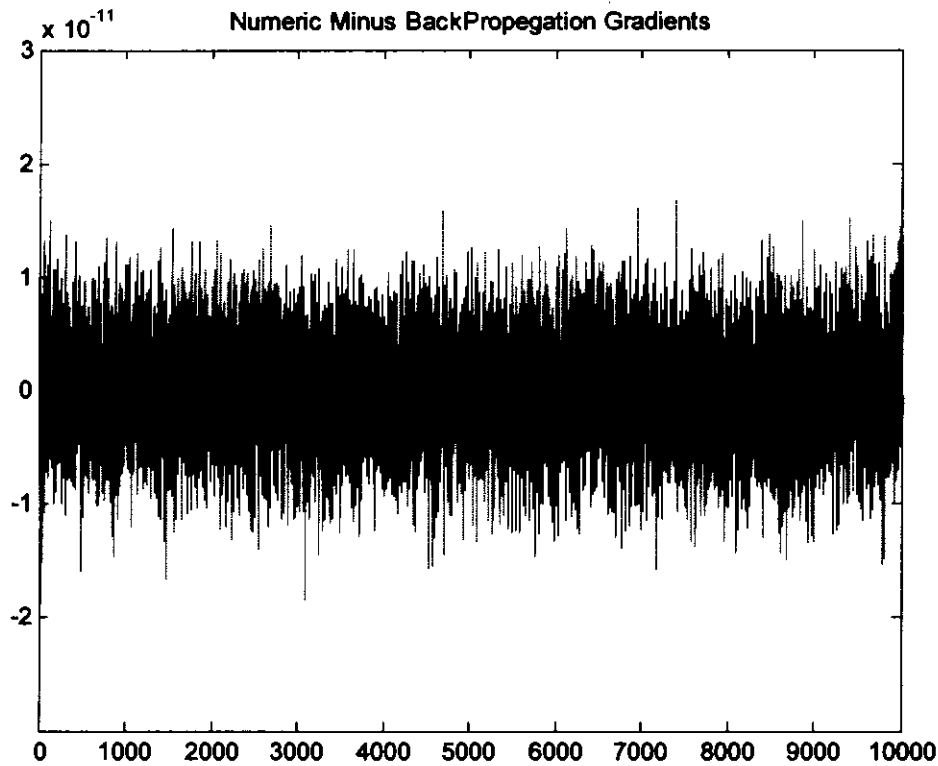
7	0	7	5	8	5	0	5	7	8
0	6	8	7	1	2	2	4	1	8
0	1	9	7	7	4	3	7	1	4
0	3	7	5	4	9	7	1	9	9
0	9	9	9	6	6	2	1	7	3
4	9	6	9	9	0	7	7	6	3
0	2	4	2	4	6	1	7	9	3
4	0	5	2	3	5	7	8	0	3
7	9	4	1	9	9	1	5	7	5
8	0	1	9	3	6	1	9	1	6

הפיקא שטעט נעמט דאסן רעזולט אומן 5000 רעזולוצאן
 העטער דאס און העמערט דא שטעט און העמערט "עם
 אולסטער אומן

figure 3

k2 pin
(checking gradient in)

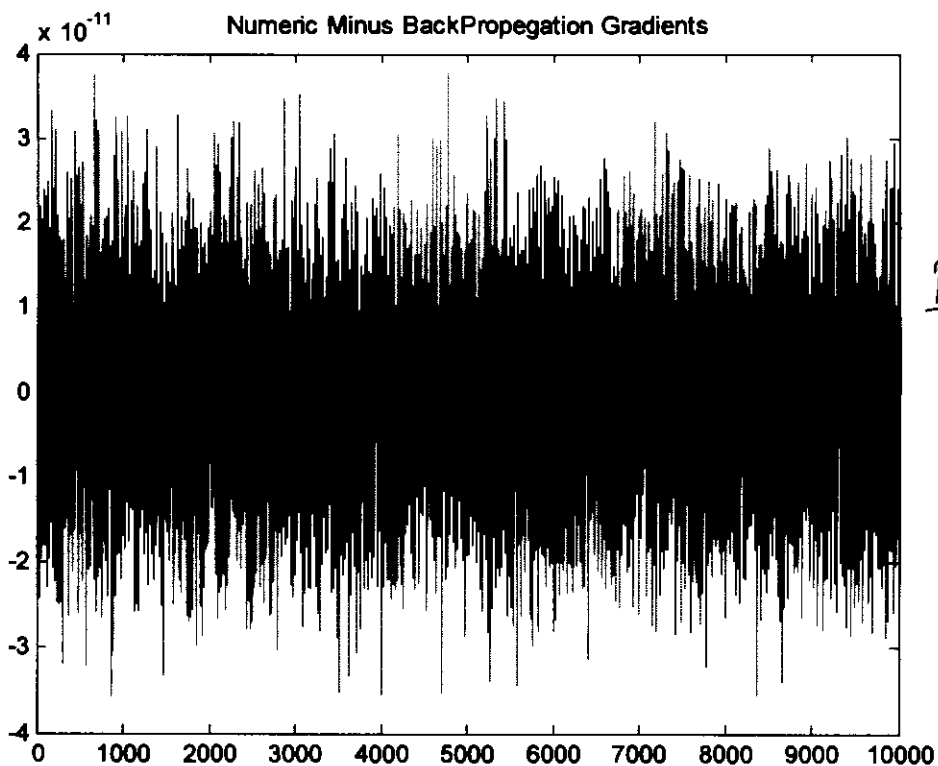
were min of wss
error, BP.



$$\eta = 0$$

Relative diff:
 $6.54 \cdot 10^{-11}$

were min of wss
error, BP.



$$\eta = 3$$

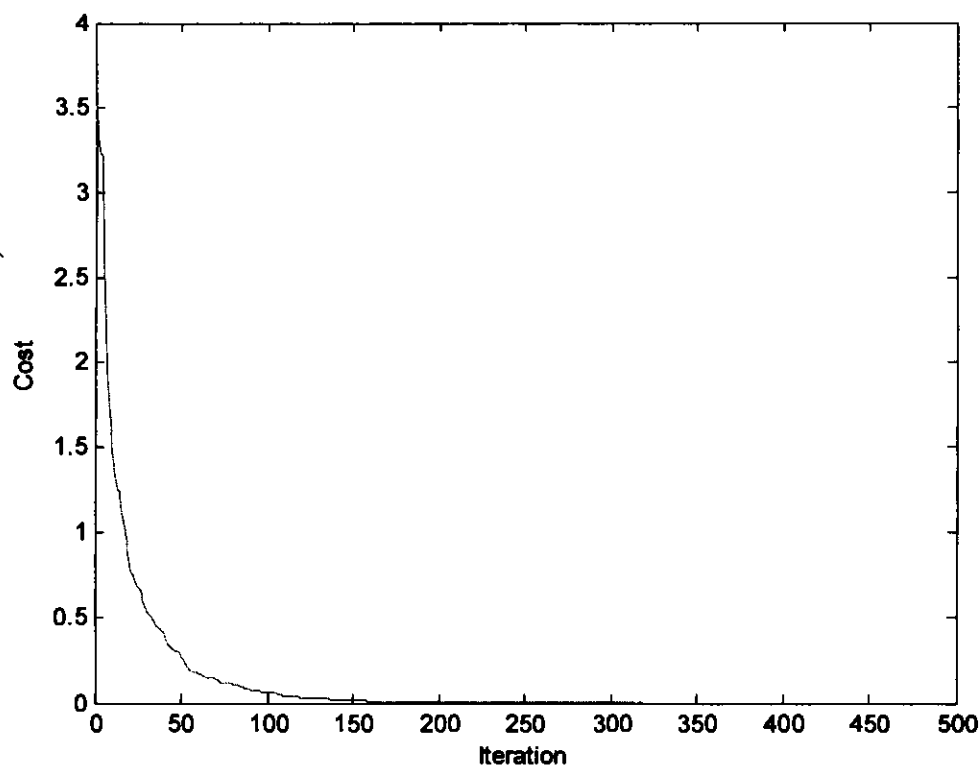
Relative diff:
 $9.82 \cdot 10^{-11}$

* הערך הממוצע של הממוצע הממוצע וכן (המקרה) מסלול של
הרבה למדנה (קט) באג מחזור בחישה (מתי...). קט מקרה ראש
שחצק הוא קטן (בזר גלל ל 10^{-11}) אגור ס=ק ג=3.

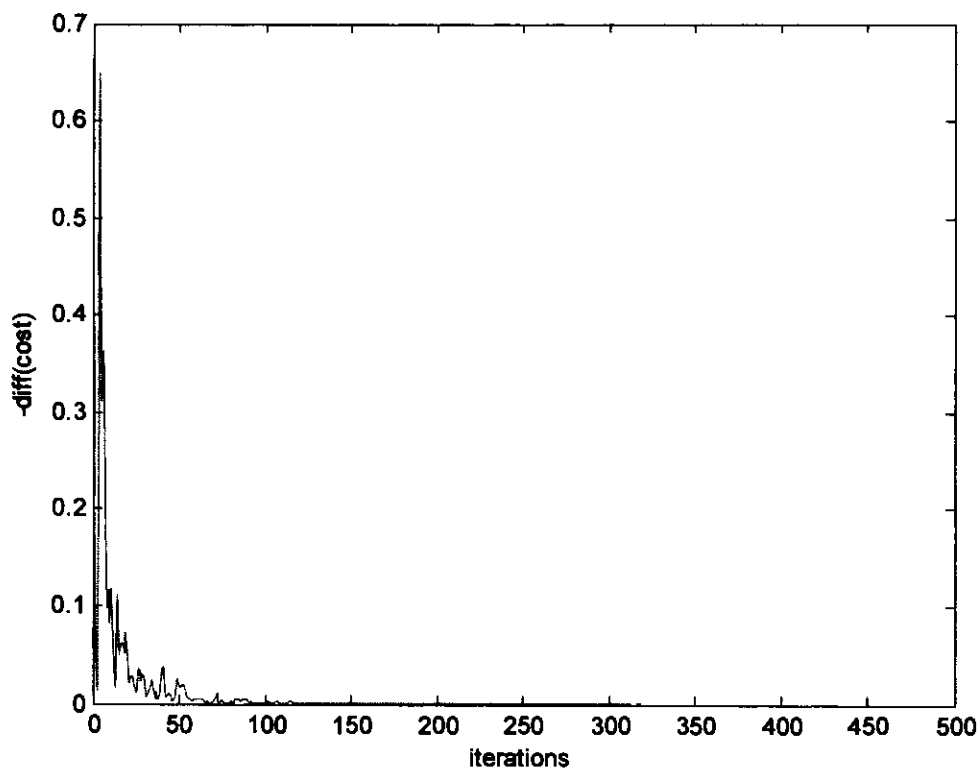
$\gamma=0$ No Regularization

תק 2

הפונקציה מוגדרת



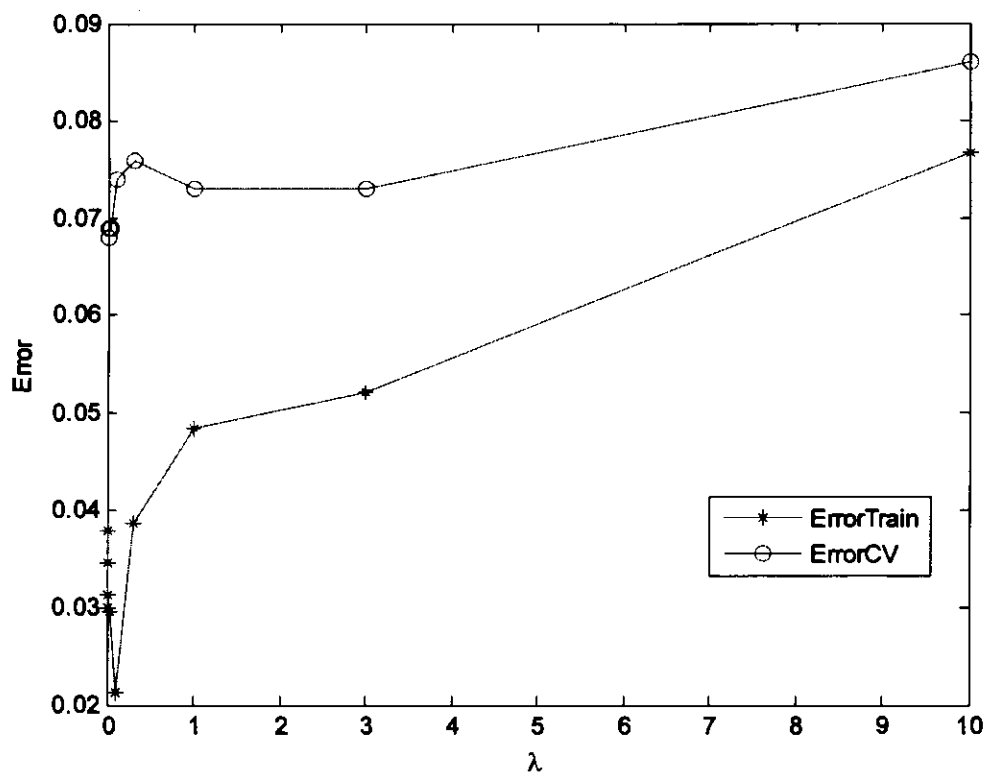
השינוי ב
הפונקציה המוגדרת



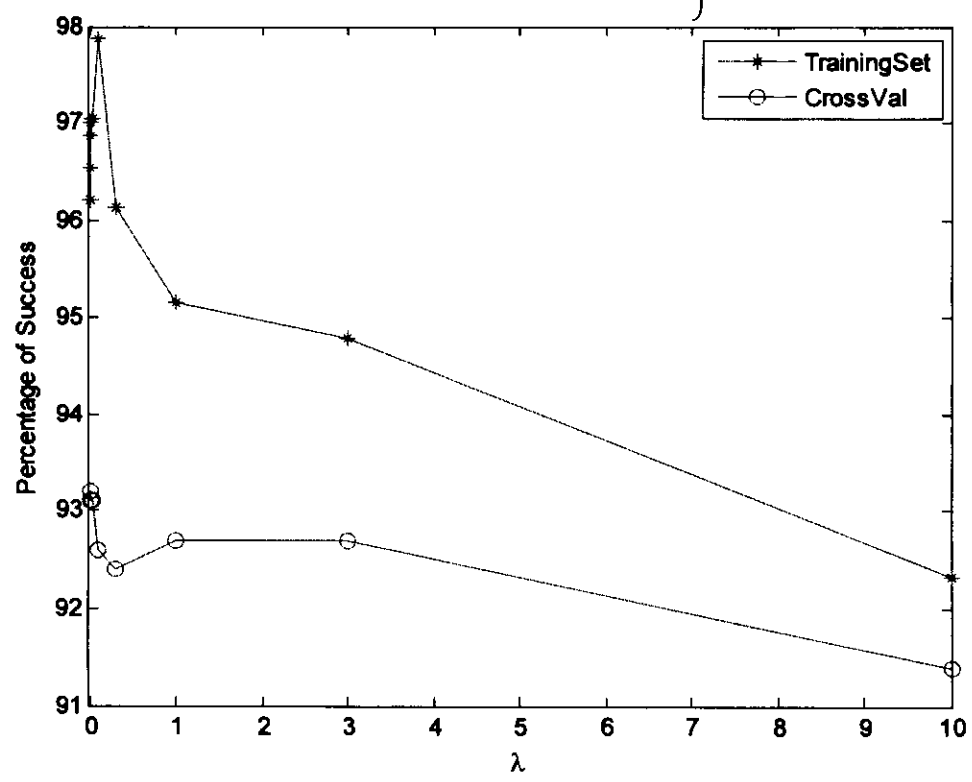
* לא מנה וקדם מהו מספר האיטרציות האופטימלי צורת ג
אם השינוי בפונקציה פוטנציאלי. כאשר שינוי זה קטן מספיק זה אומר
שכאז איטרציות לא יועילו עוד להתקדמות.
לכאז מספר איטרציות בין 50 ל-100 ייתן תוצאה טובה למדי מספיק

חלק ב'

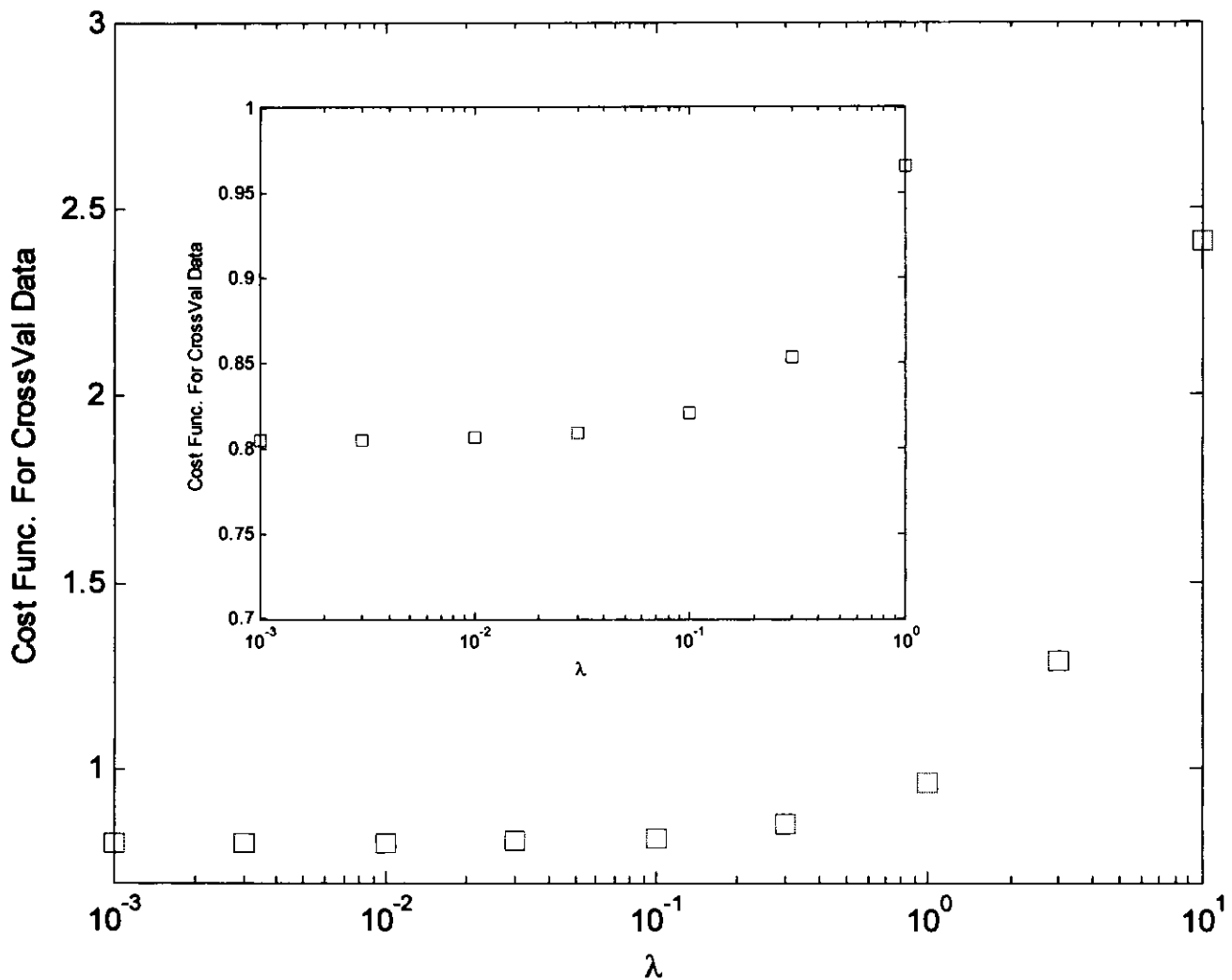
חלק יחסי של שגיאה פונקציה של λ



אחוז הצלחה פונקציה של λ



פונקציות ~~הכנסות~~ (הפסד) Θ של λ (train.set) ו-CV



ניתן לראות שעבור λ קטן מספיק (עושה קטן λ)
 ה cost func. יורד אבל עבור λ קטן מדי (אין "תשואה" יותר מדי גבוה)

כאשר $\lambda > 1$ מקבלים שגיאה מאוד גבוהה.
 (יותר מ-5% שגיאה)

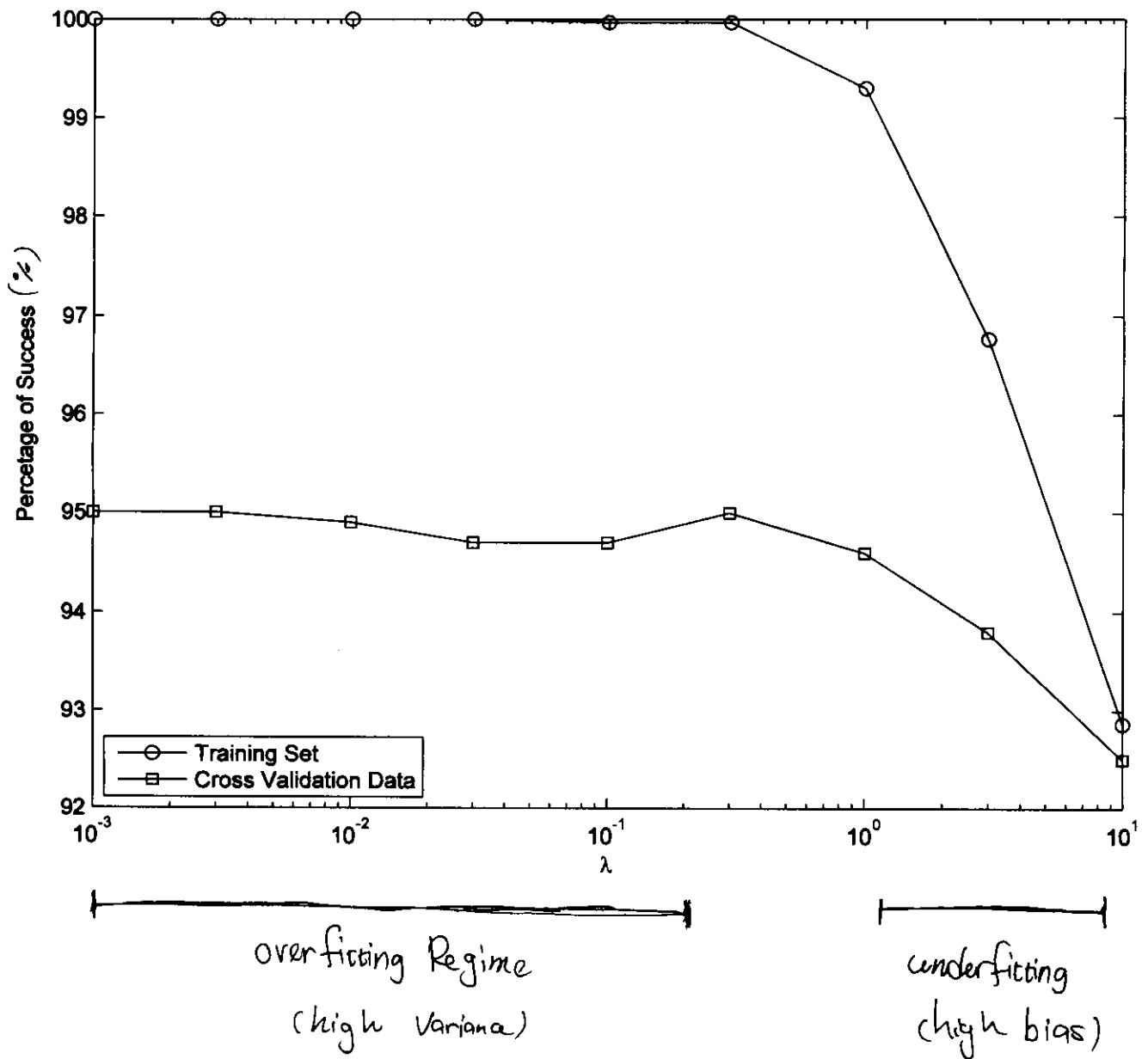
עבור $\lambda = 0.1$ נראה שיש λ overfitting כלומר השגיאה

ה training set מ-0.25 סה"כ אף עבור CV (השגיאה גבוהה (7.5%))

של האינסון (קטן) $\lambda = 1$.

optimal regularization parameter λ :

Max Iter = 200



חלק 3

Support Vector Machine.

מה שמדע נעלה בחלק זה הוא להשתמש בעקרון

one vs. all

כאשר במקום ה logistic func. הנל נשתמש במה שהקצרט עבור ה SVM (זה טיפ לט "מחזור סתחון" למידה). האלגוריתם יהיה דבר זה:

* נתן לט (הקצרט) סוקציה SVM.

* נשתמש בפוקציה \min_{cost} (או משהו דומה)

א טיפ למצא את ה- θ האידיאלי.

* עכשיו עבור S אחר מהקצרטאור נתחיל לאצט cluster

הוא שיתן עס' העקרון של one vs. all.