תרגיל בית מספר 2

מבוא ללמידה חישובית

בנימין אסתרליס ותומר סמרה

315636761 321164923
tomersamara@mail.tau.ac.il esterlis@mail.tau.ac.il

הכתובת של קבצי הקוד+ התמונות בצבע: specific/a/home/cc/students/math/esterlis/Downloads/ML-HW2

1 חלק תיאורטי

1. (א) נרצה להראות שניתן להסיר את התנאי $\xi_i < 0$ כאשר מחפשים פתרון לבעיה :

$$\min_{\mathbf{w},b,\boldsymbol{\xi}} \frac{1}{2}\mathbf{w}^T\mathbf{w} + \frac{C}{2}\sum_{i=1}^{m}\xi_i^2$$
$$s.t.\ y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i$$

ראשית נבחין שכאשר $\xi_i < 0$ מתקיים ש $\sum_{i=1}^{m}\xi_i^2 = \sum_{i=1}^{m}|\xi_i|^2$ ולכן הסכום לא משתנה כאשר ξ_i שליליים,
כעת נבחין שהמקרה שבו $y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i$ עבור $\xi_i < 0$ מוכל במקרה שבו ξ_i הינם אי שליליים (מכיוון ש $1 - \xi_i \leq 1 + |\xi_i|$), ולכן המינימום עבור $\xi_i < 0$ קטן מהמינימום עבור $\xi_i \geq 0$.

(ב)

$$\mathcal{L}(w,b,\xi,\bar{\alpha}) = \frac{1}{2}\|w\|^2 + \frac{C}{2}\sum_{i=1}^{m}\xi_i^2 + \sum_{i=1}^{m}\alpha_i\left[1 - \xi_i - y_i(x_i \cdot w_i + b)\right]$$

(ג) נגזור ונשווה לאפס את הלגאנז'יאן לפי כל אחת מהקורדינאטות כפי שעשינו בתרגול ולכן:

(ד)

$$\frac{\partial \mathcal{L}}{\partial w} = w - \sum_{i=1}^{m}\alpha_i y_i x_i = 0 \qquad \rightarrow \qquad w = \sum_{i=1}^{m}\alpha_i y_i x_i$$
$$\frac{\partial \mathcal{L}}{\partial b} = -\sum_{i=1}^{m}\alpha_i y_i = 0 \qquad \rightarrow \qquad \sum_{i=1}^{m}\alpha_i y_i = 0$$
$$\frac{\partial \mathcal{L}}{\partial \xi_i} = C\xi_i - \alpha_i = 0 \qquad \rightarrow \qquad \xi_i = \frac{\alpha_i}{C}$$

$$\begin{aligned}
\mathcal{L}(\sum_{i=1}^{m}\alpha_i x_i y_i, 0, \frac{\alpha}{C}, \alpha) &= \frac{1}{2}w \cdot w + \sum_{i=1}^{m}\xi_i\left(\frac{C}{2}\xi_i - \alpha_i\right) + b\left(-\sum_{i=1}^{m}\alpha_i y_i\right) + \sum_{i=1}^{m}\alpha_i + \left(-\sum_{i=1}^{m}\alpha_i y_i x_i\right)w = \\
&= \frac{1}{2}w \cdot w + \sum_{i=1}^{m}\xi_i\left(\frac{C}{2}\xi_i - \alpha_i\right) + \sum_{i=1}^{m}\alpha_i + \left(-\sum_{i=1}^{m}\alpha_i y_i x_i\right)w = \\
&= \frac{1}{2}\left(\sum_{i=1}^{m}\alpha_i y_i x_i\right)^2 + \sum_{i=1}^{m}\xi_i\left(\frac{C}{2}\xi_i - \alpha_i\right) + \sum_{i=1}^{m}\alpha_i - \left(\sum_{i=1}^{m}\alpha_i y_i x_i\right)\left(\sum_{i=1}^{m}\alpha_i y_i x_i\right) = \\
&= -\frac{1}{2}\left(\sum_{i=1}^{m}\alpha_i y_i x_i\right)^2 + \sum_{i=1}^{m}\xi_i\left(\frac{C}{2}\xi_i - \alpha_i\right) + \sum_{i=1}^{m}\alpha_i = \\
&= -\frac{1}{2}\left(\sum_{i=1}^{m}\alpha_i y_i x_i\right)^2 + \sum_{i=1}^{m}\frac{\alpha_i}{C}{}_i\left(\frac{C}{2} \cdot \frac{a_i}{C} - \alpha_i\right) + \sum_{i=1}^{m}\alpha_{i=} \\
&= -\frac{1}{2}\left(\sum_{i=1}^{m}\alpha_i y_i x_i\right)^2 + \sum_{i=1}^{m}\left(-\frac{a_i^2}{2C} + \alpha_i\right)
\end{aligned}$$

(ה) כעת נראה את הבעיה הדואלית :

$$\max_{\alpha}\mathcal{L}\left(\sum_{i=1}^{m}\alpha_i y_i x_i, 0, \frac{\alpha}{C}, \alpha\right) = \min\left[\frac{1}{2}(\alpha_i y_i x_i)^2 - \sum_{i=1}^{m}(\frac{a_i^2}{2C} - \alpha_i)\right]$$

2. (א) בהנחה שהבעיה הפרימלית שווה לבעיה הדואלית באופטימום, כלומר עבור $\mathbf{w}^\star, \boldsymbol{\xi}^\star, b^\star$ מתקבל המינימום בבעיה הפרימלית והוא שווה למקסימום של הבעיה הדואלית, מתקבל:

$$\frac{1}{2}\|w^\star\|^2 + C\sum_{i=1}^{m}\xi_i^\star = -\frac{1}{2}\|w^\star\| + \sum_{i=1}^{m}\alpha_i^\star$$

צד שמאל של המשואה מתקבל מה בעיה הפרימלית , צד ימין של המשוואה מתקבל מה בעיה הדואלית

(ב) ניתן לראות שמתקיים:

$$\|w^\star\|^2 = -C\sum_{i=1}^{m}\xi_i^\star + \sum_{i=1}^{m}\alpha_i^\star = \sum_{i=1}^{m}(\alpha_i^\star - C\xi_i^\star) \overset{0\le\alpha_i\le C}{\le} C\sum_{i=1}^{m}1-\xi_i^\star \overset{\xi_i\ge 0}{\le} C\sum_{i=1}^{m}1 = C\cdot m$$

ולכן נקבל שנתקיים $\|w^\star\|^2 \le C\cdot m$ ולכן מתקבל $\|w^\star\| \le \sqrt{C\cdot m}$.

3. (א) **(לא להגשה)** ראשית נבחין שהלגרנז'יאן הוא: $\mathcal{L}(w,\xi,\alpha) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{m}\alpha_i\left[1-\xi_i - y_i\cdot x\cdot w_i\right]$ וכן מתקיים של $w = \sum_{i=1}^{m}\alpha_i y_i x_i$ ולכן

$$g(\alpha) = \min_{\alpha}\mathcal{L}(w,\xi,\alpha) = \sum_i \alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j x_i x_j$$

לא נדרוש שהתקיים $\sum\alpha_i y_i = 0$ מכיוון שאין לנו ביטוי שתלוי ב b שנרצה לאפס. ולכן הבעיה הדואלית שווה לבעיה הדואלית מהכיתה רק שאנו לא דורשים שהתקיים $\sum\alpha_i y_i = 0$.

(ב)

$$\max_{\alpha_1}\sum_i\alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j x_i x_j$$
$$s.t.\ \ 0\le\alpha_i\le C$$

(ג) נגדיר את הפונקציה $f(\alpha_1) = \sum_i\alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j x_i$ ונרצה להביא אותה לאופטימום, לשם כך נרשום אותה בצורה הבאה :

$$f(\alpha_1) = \sum_i\alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j x_i x_j =$$
$$= \alpha_1 + \sum_{i=2}\alpha_i - \frac{1}{2}\alpha_1^2 y_1^2 x_1^2 - \frac{1}{2}\cdot 2\sum_{j=2}^{m}\alpha_1\alpha_j y_1 y_j x_1 x_j + \frac{1}{2}\sum_{\substack{i\ne 1\\ j\ne 1}}\alpha_i\alpha_j y_i y_j x_i x_j$$

כעת נגזור את f לפי α_1 ונקבל שמתקיים :

$$\frac{\partial f(\alpha_1)}{\partial\alpha_1} = 1 - \alpha_1 y_1^2 x_1^2 - \sum_{j=2}^{m}\alpha_j y_1 y_j x_1 x_j = 0$$

$$\alpha_1 = \frac{1-\sum_{j=2}^{m}\alpha_j y_1 y_j x_1 x_j}{y_1^2 x_1^2} \underset{y_1=\pm 1}{=} \frac{1-\sum_{j=2}^{m}\alpha_j y_1 y_j x_1 x_j}{||x_1||^2}$$

ובגלל האילוצים $0\le\alpha_i\le C$ נקבל כי:

$$\alpha_1^\star = \begin{cases} 0 & \alpha_1 < 0 \\ \alpha_1 & 0<\alpha_1<C \\ C & \alpha_1 > C\end{cases}$$

(ד) i. **סיבוכיות הזמן:** בכל איטרציה נרצה לחשב את $\alpha_1^\star = \frac{1-\sum_{j=2}^{m}\alpha_j y_1 y_j x_1 x_j}{y_1^2 x_1^2}$, ניתן לחשב זאת עלי ידי חישוב הסכום $\sum_{j=2}^{m}\alpha_j y_1 y_j x_1 x_j$ בזמן $O(d\cdot m)$ כאשר d הינו גודל המימד.
מכיוון שמכפלה פנימית של x_1 עם x_i לוקחת $O(d)$ ומבצעים את זה $O(m)$ פעמים.
ולכן לחשב את $\alpha_1^\star$ לוקח $O(m\cdot d)$.

ii. **סיבוכיות המקום:** נניח שאנו מקבלים את כל הוקטורים כנתון.
בכל איטרציה נוכל לחשב את $\alpha_1^\star$ על ידי שימוש ב $O(d)$ מקום ומכיוון שאנו מחשבים את $\alpha_1^\star$ בצורה איטרטיבית נוכל תמיד להשתמש ב $O(d)$ מקום.
אם צריך לשמור את הוקטורים סיבוכיות המקום היא $O(m\cdot d)$

4. (את השאלות בתרגיל זה ניתן לפתור גם על ידי שימוש במכפלות פנימיות)

(א) נפריך את הטענה בעזרת דוגמה נגדית:
מכוון ש K_1 ו K_2 הינם גרעינים המוגדרים חיובים אז ניתן לקחת את הגרעינים K_1,K_2 כך שמתקיים: $K_1=6I$,$K_2=5I$ (כאשר I הינו וקטור היחידה) ולכן נקבל כי:

$$K=K_1-K_2=5I-6I=-I$$

וקל לראות שהמטריצה $-I$ אינה מוגדרת חיובית, ולכן נקבל ש K אינה מוגדרת חיובית.

(ב) נראה כי הטענה אינה נכונה בעזרת דוגמא נגדית:
נבחר $\forall x,z\ K_2(x,z):=(x,z)$ ואת K_1 באופן שרירותי ולכן נקבל כי $\forall x,z\ K(x,z)=\frac{K_1(x,z)}{0}$ ומכוון שחלוקה ב0 איננה מוגדרת אזי גם K אינו מוגדר ולכן בפרט K אינו גרעיון מוגדר חיובית.

(ג) ידוע לנו ש $K=aK_1$ ולכן מכיוון ש K_1 מוגדרת חיובית לכל וקטור v מתקיים $v^T\cdot K_1\cdot v>0$
ולכן מתקיים $v^T\cdot K\cdot v=v^T\cdot aK_1\cdot v>0$ קיבלנו ש K מוגדרת חיובית.

(ד) נבחר את K_1 להיות $K_1(x,z)=1$ $\forall x,z$ לכן מתקיים כי עבור $a>0$ ש $K(x,z)=-aK_1(x,z)<0$ לכן K אינו מוגדר חיובית.

(ה) לפי מסקנה שראינו בהרצאה כי $K(x,z)$ המוגדרת ע"י $\phi(x)\cdot\phi(z)$ הינה מוגדרת חיובית ולכן עבור $\phi(x)=f(x)$ מתקבל כי K הינו גרעין מוגדר חיובית.

(ו) ראשית, ניזכר שראינו בתרגול כי מכפלה של גרעינים היא גרעין מוגדר חיובית ולכן K^nהינו גרעין מוגדר חיובית לכל K ו $n>0$.
בנוסף ראינו בתרגול שסכום של גרעינים מוגדרים חיובית הוא גרעין מוגדר חיובית ולכן מכיוון ש p הינו פולינום עם מקדמים חיוביים נקבל בברור כי Kהינו גרעין מוגדר חיובית.

5. (א) נוכיח את הטענה באינדוקציה על T:
<u>בסיס:</u> $(T=1)$
ניתן לראות כי אם ישנה שגיאה בחיזוי, ומכוון שאנו באיטרציה הראשונה נקבל כי $w_1=\pm x_1$ ואם אין שגיאה בחיזוי נקבל כי $w_1=\bar{0}$ וכך או כך נקבל כי w_1 תלוי ליניארית ב x_1.
<u>צעד האינדוקיצה:</u> $(T\rightarrow T+1)$
נניח כי מתקיים עבור T ונוכיח עבור $T+1$.
ניתן לראות כי אם ישנה טעות שוב נקבל כי $w_{T+1}=w_T\ \pm\ x_T$ ולכן מתקבל כי w_{T+1}הינו תלוי ב x_T וב w_T, כעת לפי הנחת האינדוקציה אנו יודעים כי w_T הינו תלוי לינארית ב $\{x_i\}_{i=1}^{T-1}$,
ולכן נקבל כי w_{T+1} גם כן תלוי ליניארית ב $\{x_i\}_{i=1}^{T}$.

(ב) ראשית ידוע כי לא ניתן לשמור את המסווג w מכוון ש ϕ אשר מקבלים יכולה להיות ממימד ∞-י ולא ניתן לשמור משהו בזיכרון ∞-י.
ולכן נעזר במה שהראנו בסעיף (a)אשר אומר כי ה $w_T=\sum_{i=0}^{T-1}\alpha_i\cdot x_i$ עבור α_iהמתאים ל x_i.
כעת לפי החישוב של סעיף (a) ניתן לראות כי $\alpha_i=m_i\cdot y_i$ כאשר:

$$m_i - no.\,of\,times\ x_i\ was\ classified\ wrong$$
$$y_i - the\,true\,label\,of\,x_i$$

כעת ניתן לראות כי החישוב באלגוריתם המקורי נעשה באופן הבא:

$$\langle w_T, x\rangle = \left\langle \sum_{i=1}^{T-1} \alpha_i x_i, x \right\rangle = \sum_{i=1}^{T-1} \alpha_i \langle x_i, x\rangle = \sum_{i=1}^{T-1} m_i \cdot y_i \langle x_i, x\rangle$$

ולכן כאשר נרצה להפעיל את ϕ נקבל כי:

$$\langle w_T, x\rangle = \left\langle \sum_{i=1}^{T-1} \alpha_i \phi(x_i), \phi(x) \right\rangle = \sum_{i=1}^{T-1} m_i \cdot y_i \langle \phi(x_i), \phi(x)\rangle$$

כעת נתאר את האלגוריתם:
בהינתן $\{x_i\}_{i=1}^m$ -דגימות, $\{y_i\}_{i=1}^m$-הסיווג הנכון של x_i בהתאמה ו K להיות הגרעין המתקבל

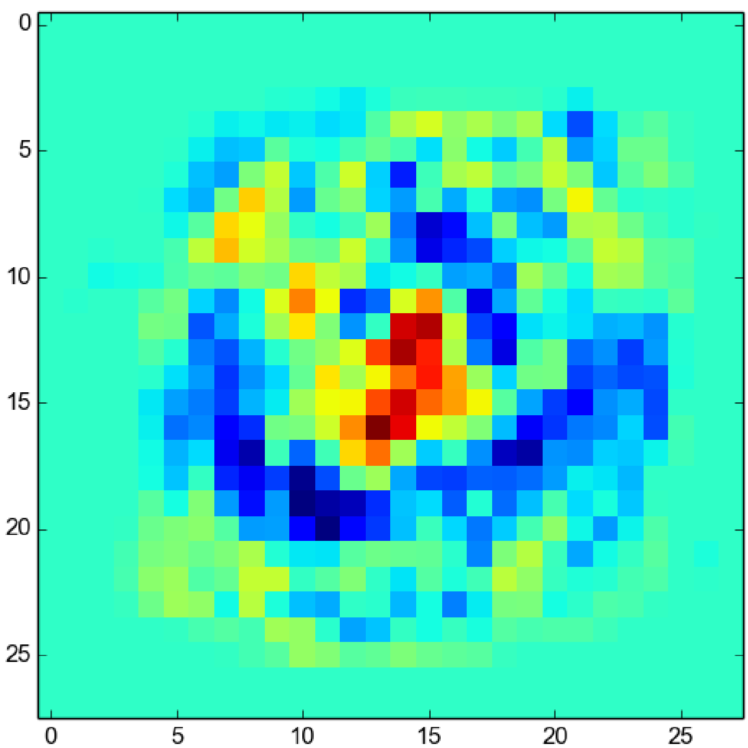
- נגידר את המסווג שלנו להיות C אשר יאותחל להיות וקטור אפסים ממימד m.
- $\forall x_k \in \{x_i\}_{i=1}^m$.:
  - נחזיר את התוצאה של $\sum_{i=1}^m y_i \cdot C[k] \cdot K(x_i, x) \leq 0$
  - כעת אם התוצאה שהחזרנו שונה מ y_i נגדיל את הערך שב $C[k]$ב1.

2 חלק מעשי

1. (א) ניתן לראות כי הטבלה שיצאה הינה:

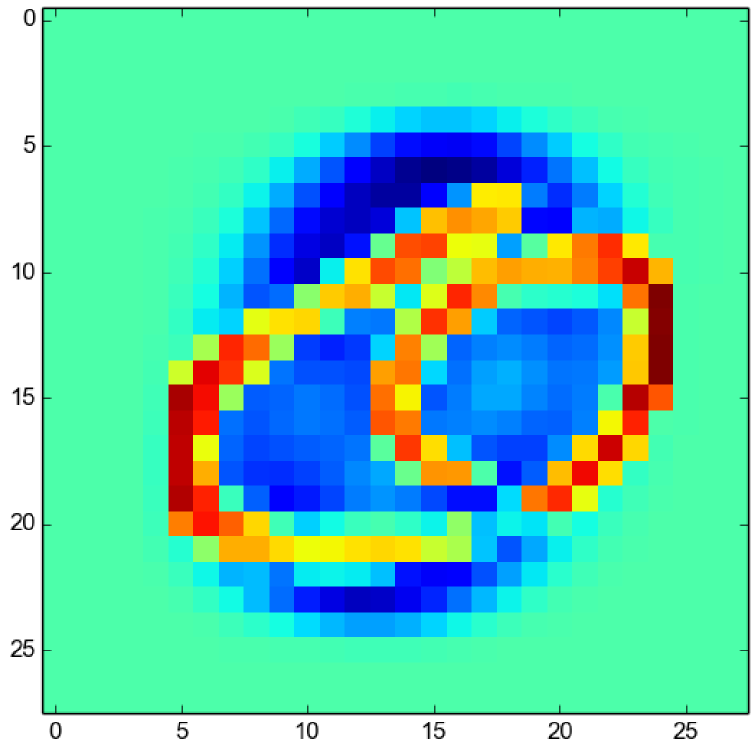
$mean$	%95	5%	n
0.853254861822	0.905322415558	0.67451381781	5
0.911243602866	0.948311156602	0.874104401228	10
0.954964176049	0.980578300921	0.913868986694	50
0.96818321392	0.983137154555	0.948285568065	100
0.980061412487	0.989764585466	0.96389457523	500
0.984636642784	0.990813715455	0.973285568065	1000
0.988316274309	0.992860798362	0.979017400205	5000

(ב) ההבדל העיקרי בין שמיניות לאפסים הוא שבשמיניות הפיקסלים הנמצאים במרכז שחורים לעמות אפסים אשר בהם הפיקסלים באמצע לבנים.$[q1\_b()]$

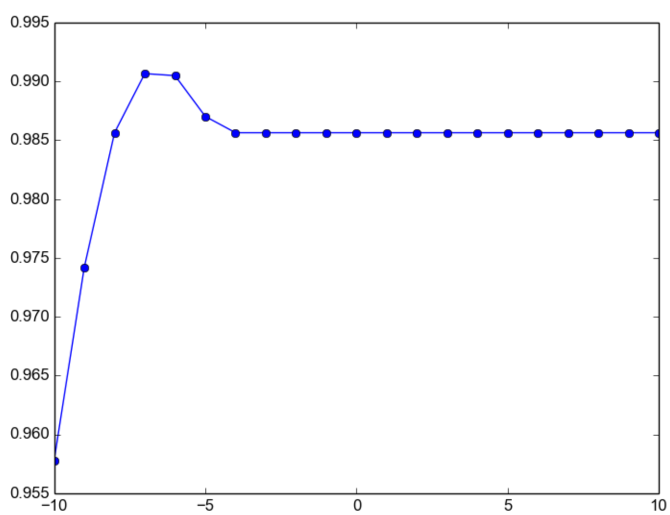


(ג) אחוזי הדיוק הוא:0.987717502559

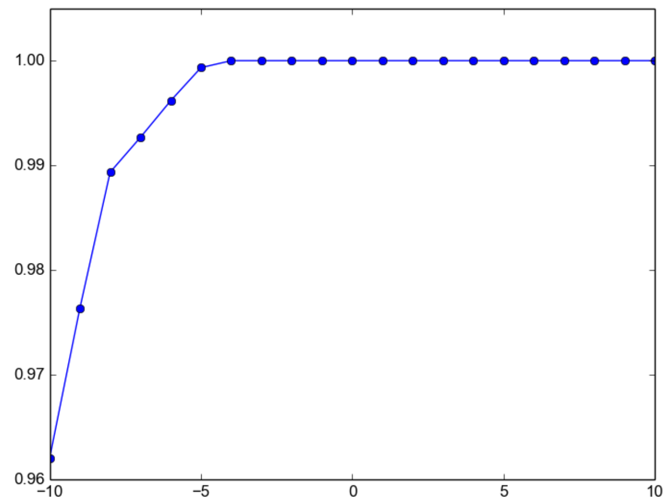
(ד) ניתן לראות כי האלגורים שלנו טעה לזהות את המספר הזה מכוון שלא ניתן לזהות בוודאות אם המספר שלנו הוא אפס או שמונה אשר שוכב, שכן הפיקסלים המשמועתיים דומים .



2. (א) i. הגרף מייצג את הדיוק על $validation\ data$ כפונקציה של $\log C$



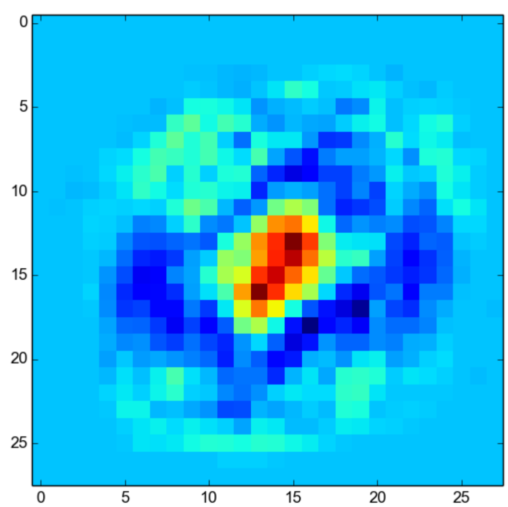
ii. הגרף מייצג את הדיוק על $train\ data$ כפונקציה של $\log C$



וניתן לראות כי ה C הטוב ביותר שהתקבל הינו: $C = 10^{-7}$.

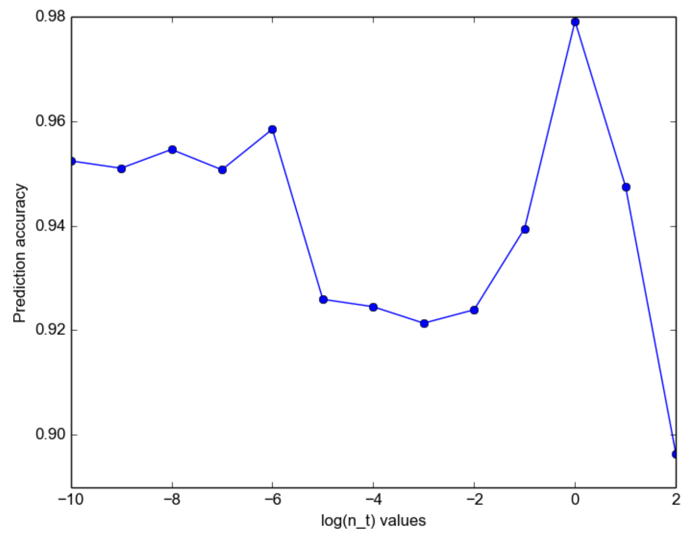
(ב) אנו יכולים לראות כי ה $training\ error$קטן ככל ש C גדל , כי C הוא המקדם של השגיאה, משמע ככל שזה גדל אז לכל סיווג השגיאב תגדל גם כן. ולכן ככל ש Cגדל כך מספר הטעויות קטן.

(ג) התמונה שהתקבלה :

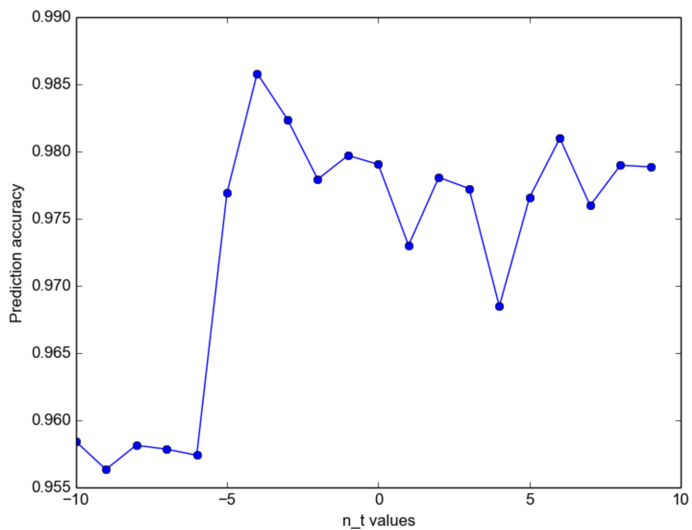


(ד) הדיוק הוא: 0.990647731209 והוא מתקבל עבור $C = 10^{-7}$.

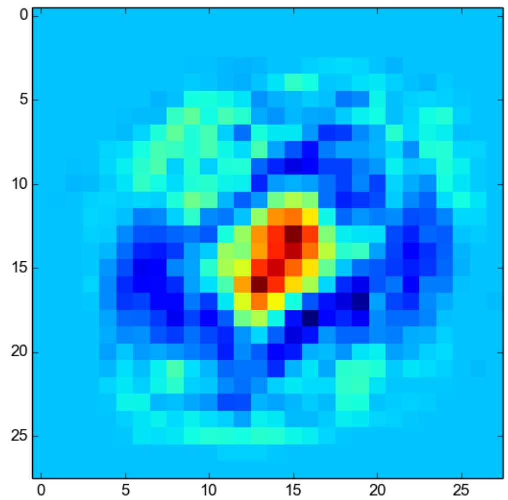
3. (א) ה η_0 הטוב ביותר הוא: 1 והוא מתקבל כתוצאה מהגרף:



(ב) ה C הטוב ביותר הוא: 10^{-4} והוא מתקבל כתוצאה מהגרף:



(ג) ה w המתקבל הוא:



(ד) הדיוק שהתקבל הינו 0.991811668373