מבוא ללמידה חישובית – תרגיל בית 3

שם : דורי רימון

ת"ז : 323996843

שאלות תיאורטיות:

שאלה 1:

שאלה 1

יהי $0 < \gamma < 1$. נגדיר $m = d = \lfloor \frac{1}{\gamma^2} \rfloor$ ו-

$\{x_i\}_{i=1}^{m}$ להיות הבסיס הסטנדרטי של $\mathbb{R}^d$. בנוסף נגדיר

$y_i = -1$, $\forall 1 \le i \le m$ (כלומר כל הנקודות מקוטלגות שלילית).

נניח

$$(b) \quad \frac{y_i \cdot x_i \cdot w_i^*}{\|w^*\|} \ge \gamma$$

$$\Rightarrow \frac{(0, \dots, 0, \overset{i}{-1}, 0, \dots, 0) \cdot w^*}{\|w^*\|} \ge \gamma$$

$$\Rightarrow -w_i \ge \|w^*\| \cdot \gamma$$

ה

$$\Rightarrow w_i \le -\|w^*\| \cdot \gamma$$

נסמן

$w^* = (w_1, \dots, w_d)$

נבחר לנרמל הנורמה את w^* כך ש-

$w_1 = w_2 = \dots = w_d$.

ולכן אנו מקבלים ש

$$\Rightarrow w_i \le -\sqrt{d \cdot w_i^2} \cdot \gamma$$

$$\Rightarrow w_i \le -\sqrt{d} \cdot |w_i| \cdot \gamma$$

$$\Rightarrow \frac{w_i}{|w_i|} \le -1$$

$$\Rightarrow w_i = -1$$

פתרון אפשרי

וקטור שלילי $w^* = (-1, \ldots, -1)$. הוקטור מפריד (נורמל) לא פם כן נורמה

ב-w^* יתקבל את התיוג הנכון (וקטור זה — ו-(-1)).

וקטור מתקיים $\forall 1 \le t \le m,\ sign(w^* \cdot x_t) = -1$.

עכשיו נראה, שלאחר שהגדרנו את D המערך הדרוש, להראות — כי 3 המשוואות מתקיימות:

(a) $\|x_i\| = 1$

מתקיים כיוון שאלו וקטורי הבסיס הסטנדרטי.

(b) $\frac{y_i x_i w^*}{\|w^*\|} \ge \gamma$

בחרנו את w^* כך שהמשוואה זו תתקיים.

(c) $M = \lfloor \frac{1}{\gamma^2} \rfloor$

נוכיח כי הפרספטרון יטעה בכל שלב, באופן אינדוקטיבי:

$t = 1$: $w_1 = 0 \Rightarrow \tilde{y}_t = sign(w_t \cdot x_t) =$
$= sign(\bar{0}) = +1 \neq -1$

צעד האינדוקציה

$t \to t+1$: mistake at stage $t \Rightarrow w_{t+1} = w_t + y_t x_t$
$= w_t - x_t =$
$= (\underset{0}{-1}, \underset{1}{-1}, \ldots, \underset{t}{-1}, \underset{t+1}{0}, 0, \ldots, \underset{d}{0})$

$\Rightarrow \tilde{y}_{t+1} = sign(w_{t+1} \cdot x_{t+1}) = sign(\bar{0}) = +1 \neq -1$

ואכן יש טעות בכל שלב, כלומר m טעויות, כלומר $\lfloor \frac{1}{\gamma^2} \rfloor$ טעויות כנדרש.

שאלה 2:

שאלה 2

ראשית, נשים לב כי $|H| = \binom{d}{2}$. יהי $1 \le a \le d$.

$$|H_a^1| := |\{h \in H \mid h(a) = 1\}| =$$
$$= |\{h_{i,a} \mid i < a\} \cup \{h_{a,j} \mid a < j\}| = d-1$$
$$\Rightarrow |H_a^0| := |\{h \in H \mid h(a) = 0\}| = |H| - |H_a^1| =$$
$$= \binom{d}{2} - (d-1) = \frac{(d-1)(d-2)}{2}$$

מהנתון $d \ge 6$, נקבל $|H_a^1| < |H_a^0|$ ולכן עד השגיאה הראשונה האלגוריתם יחזה תמיד $y=0$.

אם כן בבחירת השגיאה הראשונה a^*, מתקיים $y_t = 1$ (כאשר השגיאה מתרחשת באיטרציה t).

כעת, לפי נוסחת האלגוריתם, $H_{t+1} = \{h \in H \mid h(a^*) = 1\} =$
$$= \{h_{i,a^*} \mid i < a^*\} \cup \{h_{a^*,j} \mid a^* < j\}$$

כעת עבור $1 \le b \le d$,

בה"כ, $a^* < b$

$$|H_b^1| = |\{h \in H_{t+1} \mid h(b) = 1\}| = |\{h_{a^*,b}\}| = 1$$
$$\Rightarrow |H_b^0| = |\{h \in H_{t+1} \mid h(b) = 0\}| = |H_{t+1}| - |H_b^1| =$$
$$= d-1-1 = d-2$$

ולכן עדיין, עד השגיאה השנייה האלגוריתם יחזה תמיד 0.

מכאן, בבחירת השגיאה השנייה b^*, מתקיים $y_{b^*} = 1$. כמו כן,

$$H_{b^*+1} = \{h_{a^*,b^*}\}$$

ומהנחת realizable מכאן לא יהיו שגיאות נוספות.

אם כן אכן $M(A_{HAL}, H) = 2$ כנדרש.

שאלה 3:

שאלה 3

הוכחה: נבחין כי אינטרוול הוא רציף. אם 2 נקודות נמצאות באינטרוול כלשהו — כל אחת שביניהן יהיו גם הן באינטרוול.

נסמן את הנקודות שמסווגות ע"י c כ- $x_1, x_2, \dots, x_m$, ונניח בה"כ כי הן ממוינות בסדר עולה.

אם כן עבור כל קומבינציה של הנקודה b (הקצה הימני של האינטרוול) נספור כמה אופציות יש ל-a (הקצה השמאלי).
נשים לב כי אנו מדברים במספר סופי, שכן כל מיקום של קצה אינטרוול שלא עובר נקודה לא משנה את $h_{a,b}$.

עבור קביעת הנקודה b בקצה הימני (מימין ל-x_m) יש m אופציות ל-a.
עבור $x_m > b > x_{m-1}$, יש $m-1$ אופציות עבור a.
וכך הלאה עד אשר עבור $x_2 > b > x_1$ יש אופציה אחת בודדת ל-a.

לבסוף, ישנה גם האופציה ש-a ו-b מצויים מחוץ אל הטווח. כל אלו תוצאות בנפרד יחידה ומחולקה.

$$\Rightarrow |H_c| = \underbrace{1}_{\text{הפונק' הריקה}} + \underbrace{\sum_{k=1}^{m} k}_{\text{סכימת כל האופציות שציינו קודם}} = 1 + \frac{(m+1)\cdot m}{2}$$

שאלה 4:

שאלה 4

תהי $f : \mathbb{R}^n \to \mathbb{R}$ קמורה, $A \in \mathbb{R}^{n \times n}$ ו- $b \in \mathbb{R}^n$. נגדיר

תוכיח כי $g(x) := f(Ax+b)$ היא קמורה.

יהי $0 \le t \le 1$, $x_1, x_2 \in \mathbb{R}^n$, $x_1 \neq x_2$

$$g(tx_1 + (1-t)x_2) \overset{\text{מהגדרת } g}{=} f(A(tx_1 + (1-t)x_2) + b) =$$
$$= f(t \cdot Ax_1 + (1-t) \cdot Ax_2 + (t + (1-t)) \cdot b) =$$
$$= f(t(Ax_1 + b) + (1-t) \cdot (Ax_2 + b)) \le$$
$$\underset{\text{קמורה } f}{\le} t \cdot f(Ax_1 + b) + (1-t) \cdot f(Ax_2 + b) \overset{\text{מהגדרת } g}{=} t \cdot g(x_1) + (1-t) \cdot g(x_2)$$

וקיבלנו את הדרוש - g קמורה.

שאלה 5:

שאלה 5

כפי שראינו בהרצאה, נתעלם תחילה מהאילוצים האי-שליליות של ההסתברויות.

נשתמש בכופלי לגרנז' על H תחת התנאי:

$$\sum_{i=1}^{n} p_i = 1 \Rightarrow h(p) := \sum_{i=1}^{n} p_i - 1 = 0$$

$$* \quad L(p,\lambda) = H(p) + \lambda \cdot h(p) = \sum_{i=1}^{n} (-p_i \log p_i + \lambda p_i) - \lambda$$

$$* \quad \nabla_p L(p,\lambda) = (-1 - \log p_1 + \lambda, \ldots, -1 - \log p_n + \lambda)$$

$$* \quad \nabla_\lambda L(p,\lambda) = \sum_{i=1}^{n} p_i - 1$$

$$\nabla_p L(p,\lambda) = 0 \Rightarrow \forall 1 \le i \le n, \; -1 - \log p_i + \lambda = 0$$

$$\Rightarrow \log p_i = \lambda - 1$$

$$\Rightarrow p_i = e^{\lambda - 1} \quad (*)$$

$$\nabla_\lambda L(p,\lambda) = 0 \Rightarrow \sum_{i=1}^{n} p_i - 1 = 0$$

נציב את $(*)$ $\Rightarrow \sum_{i=1}^{n} e^{\lambda-1} = 1 \Rightarrow n \cdot e^{\lambda-1} = 1$

$$\Rightarrow e^{\lambda-1} = \frac{1}{n} \Rightarrow \lambda - 1 = \log\frac{1}{n}$$

$$\Rightarrow \lambda = \log\left(\frac{1}{n}\right) + 1$$

$$\Rightarrow \forall 1 \le i \le n, \; p_i = e^{\log\left(\frac{1}{n}\right) + 1 - 1} = e^{\log\frac{1}{n}} = \frac{1}{n}$$

λ נציב חזרה ב-$(*)$

וקיבלנו את ההתפלגות האחידה.

בנוסף, כמובן, $\forall 1 \le i \le n, \; p_i \ge 0$.

שאלה 6:

שאלה 6

תהי A מטריצה PSD. $f(w) = w^T A w$. מהנתון על A נקבל ש-

לכל $w \neq \vec{0}$, $f(w) > 0$.

$$\nabla f(w) = (A + A^t)\, w \overset{A \text{ סימטרית}}{=} (A + A)\, w = 2Aw$$

$$\Rightarrow w_{t+1} = w_t - \eta \nabla f(w_t) = w_t - \eta 2 A w_t = (I - 2\eta A)\, w_t$$

$$\Rightarrow w_T = (I - 2\eta A)^{T-1} \cdot w_1$$

עבור v - ו"ע עם ע"ע λ של A, מתקיים כי v ו"ע עם ע"ע $1 - 2\eta\lambda$ של המטריצה $I - 2\eta A$

כמו כן, v ו"ע עם ע"ע $(1-2\eta\lambda)^{2T-2}\lambda$ של המטריצה $Q^T A Q$ עבור $Q = (I - 2\eta A)^{T-1}$.

כיוון ש- A סימטרית, ניתן להרכיב בסיס של $\mathbb{R}^n$ מו"ע שלה. נסמן בסיס זה כ- $\{v_i\}_{i=1}^n$.

$$\Rightarrow w_1 = \sum_{i=1}^n \alpha_i v_i$$

$$\Rightarrow f(w_T) = f(Q w_1) = (Q w_1)^{\overset{\text{transpose}}{t}} A\, (Q w_1) =$$

$$= w_1^t Q^t A Q w_1 = w_1^t Q^t A Q \left(\sum_{i=1}^n \alpha_i w_i\right) =$$

$$= \sum_{i=1}^n \alpha_i \left(w_1^t Q^t A Q v_i\right) =$$

$$= \sum_{i=1}^n \alpha_i \left(w_1^t (1 - 2\eta\lambda_i)^{2T-2} \lambda_i v_i\right) =$$

$$= \sum_{i=1}^n (1 - 2\eta\lambda_i)^{2T-2} \lambda_i \left(w_1^t \alpha_i v_i\right)$$

נדרוש שיתקיים $\sum_{i=1}^n (1 - 2\eta\lambda_i)^{2T-2} \lambda_i = \frac{\varepsilon}{2\|w_1\|^2}$

נקבל $T \leq O(\log \frac{1}{\varepsilon})$.

אם יתקיים השוויון הבא — ,

$$f(w_t) = \sum_{i=1}^{n} (1-2\eta\lambda_i)^{2T-2} \lambda_i (w_1^T \alpha_i v_i) =$$

$$= \frac{\epsilon}{2\|w_1\|^2} \cdot \sum_{i=1}^{n} w_1^T \alpha_i v_i \le \frac{\epsilon}{\|w_1\|^2} \cdot \sum_{i=1}^{n} w_1^T \alpha_i v_i = \epsilon$$

כנדרש.

נסמן $\lambda_m = \min_i \lambda_i$, $\lambda_M = \max_i \lambda_i$

אנו דורשים ש-

$$\sum_{i=1}^{n} (1-2\eta\lambda_i)^{2T-2} \lambda_i \le \lambda_M \sum_{i=1}^{n} (1-2\eta\lambda_i)^{2T-2} \le$$

$$\le \lambda_M \cdot n \cdot (1-2\eta\lambda_m)^{2T-2} = \frac{\epsilon}{2\|w_1\|^2}$$

ניקח η כך שיתקיים

$1-2\eta\lambda_M > 0$ $\quad 0 < \eta < \frac{1}{2\lambda_M}$

כך לכל $1 \le i \le n$, $1-2\eta\lambda_i > 0$ ובפרט $0 < 1-2\eta\lambda_i \le 1-2\eta\lambda_m$

$$\Rightarrow (1-2\eta\lambda_m)^{2T-2} = \frac{\epsilon}{2\lambda_M \cdot n \cdot \|w_1\|^2}$$

$$\therefore T = \frac{\log\left(\frac{\epsilon}{2\lambda_M \cdot n \cdot \|w_1\|^2}\right)}{2\log(1-2\eta\lambda_m)} + 2$$

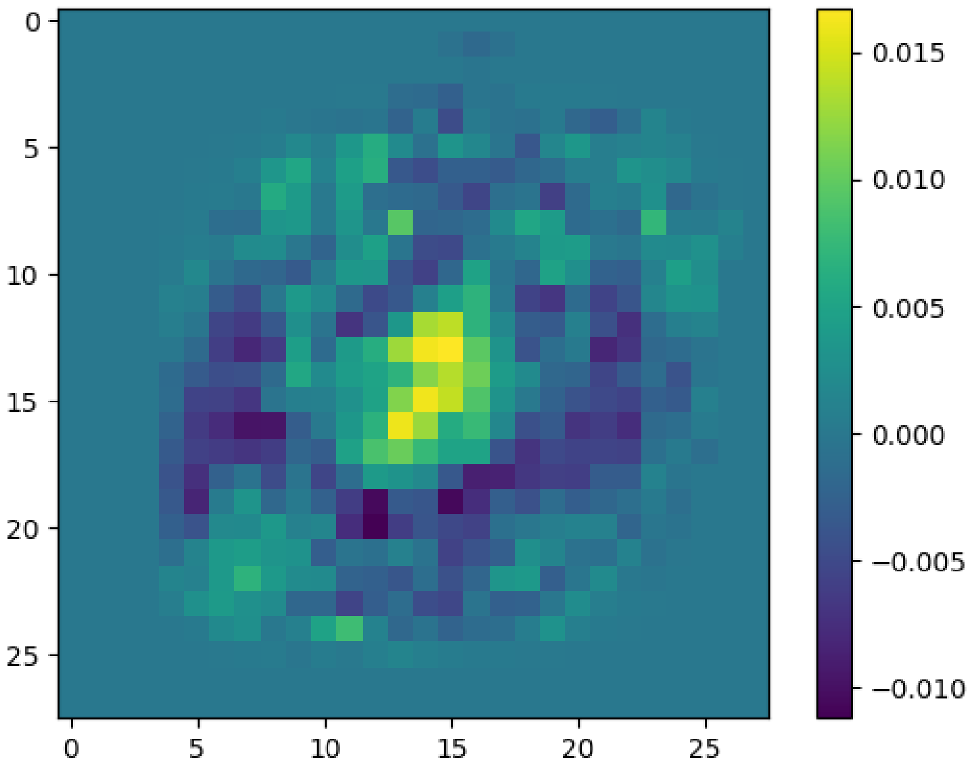
ואם נבחר $\eta = \frac{1}{4\lambda_M}$ נקבל $T = O(\log\frac{1}{\epsilon})$ כנדרש.

<u>מטלה מעשית:</u>

א.

n	mean	5%	95%
5	0.798	0.482	0.911
10	0.841	0.633	0.956
50	0.937	0.861	0.978
100	0.954	0.901	0.981
500	0.977	0.957	0.987
1000	0.984	0.968	0.990
5000	0.986	0.975	0.992

ב.



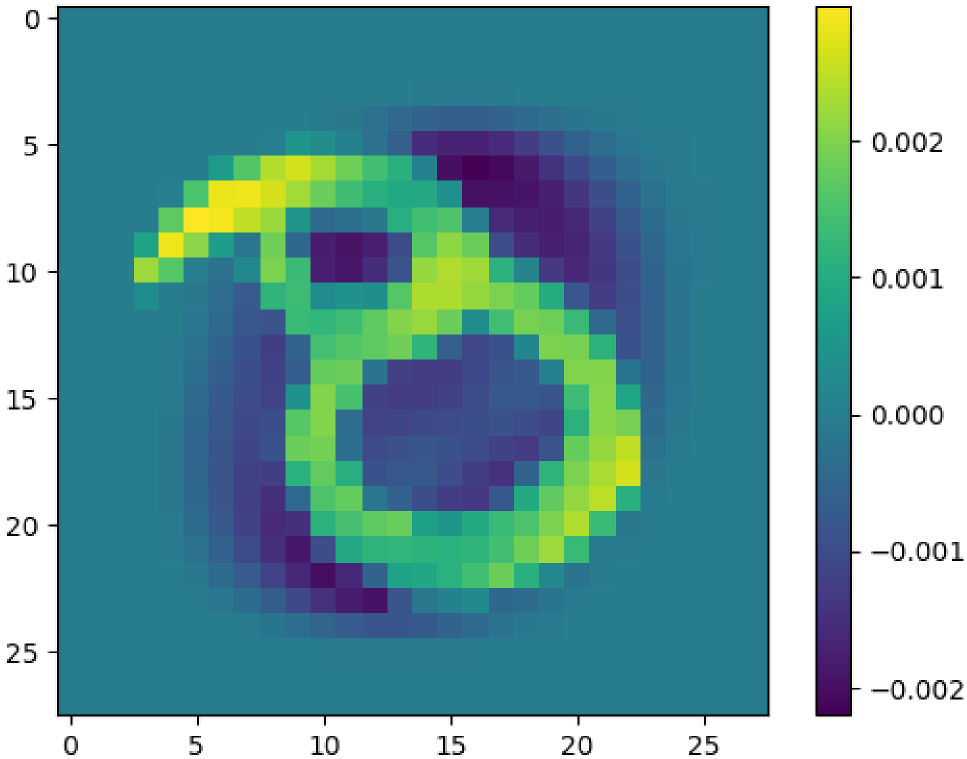
ניתן לראות כי במרכז התמונה הערכים גבוהים ובאזור שמסביב למרכז הערכים שליליים.

הסיבה לכך היא כי האזור שבמרכז התמונה הוא זה שבו נחתכים הקווים של הספרה 8, ובכך נבדלות הספרות. ע״י שיוך ערך גבוה לחלק זה אנו נותנים משמעות חיובית גדולה לפיקסלים באזור זה. לעומת זאת, האזורים השליליים המקיפים את המרכז הם אלו המבדילים בין 8 ו- 0 (רק 0 עובר בדרך כלל באזורים אלו) ולכן אנו משייכים להם משמעות שלילית.

לבסוף, נשים לב כי הערכים מסביב, שאינם רלוונטיים לא ל- 8 ולא ל- 0, מחזיקים ערכי 0, שכן אין השפעה בתיוג לאף כיוון.

ג. הדיוק הוא 0.989, כלומר 98.9%.

ד.



אנו רואים כאן תמונה של 8 שתויג כ- 0. תיוג זה מסתדר עם המסווג שראינו בסעיף ב, שכן החיתוך בין הקווים שבספרה 8 נוטה מעלה יחסית מהמרכז. כלומר במקום להיות מוכפל בערך חיובי גבוה כרצוי, נקבל בדיוק את התוצאה ההפוכה – נכפול ביטים רבים במספרים שליליים.