פתרון שיעורי בית 5 – רוני קוסיצקי 205817893

<u>שאלה 1</u>

ניעזר ברמז. מכיוון ש$x_1, \dots x_n$ שונים אזי המטריצה הנ"ל היא ממםד n.

$$\begin{pmatrix} 1 & x_1^1 & \dots & x_1^q \\ 1 & x_2^1 & \dots & x_2^q \\ \vdots & & & \\ 1 & x_n^1 & \dots & x_n^q \end{pmatrix}$$

נראה שהמטריצה הזו היא ממםד n:

$$\phi = \begin{pmatrix} \phi(x_1)^T \\ \vdots \\ \phi(x_n)^T \end{pmatrix}$$

מהטענה מהתרגול שברמז, אנו יודעים שהמידע$\phi(x_1), \dots, \phi(x_n)$ ניתן להפרדה ליניארית, ולכן אלגוריתם $hard\ SVM$ ישיג שגיאה 0. נמצא את הפונקציה $\phi(x)$ המתאימה לקרנל המתואר:

$$K(x_i, x_j) = (1 + x_i x_j')^q =_{Newton's\ binom} \sum_{m=0}^{q} \binom{q}{m} x_i^m x_j^m = \sum_{m=0}^{q} \binom{q}{m}^{\frac{1}{2}} x_i^m \cdot \binom{q}{m}^{\frac{1}{2}} x_j^m = \phi(x_i)^T \cdot \phi(x_j)$$

מכאן נקבל שכל שורה i במטריצה צריכה להיות:

$$\phi_{i'th\ row} = \phi(x_i)^T = \begin{pmatrix} 1 & \binom{q}{1}^{0.5} x_i & \binom{q}{2}^{0.5} x_i^2 \dots \binom{q}{q}^{0.5} x_i^q \end{pmatrix}$$

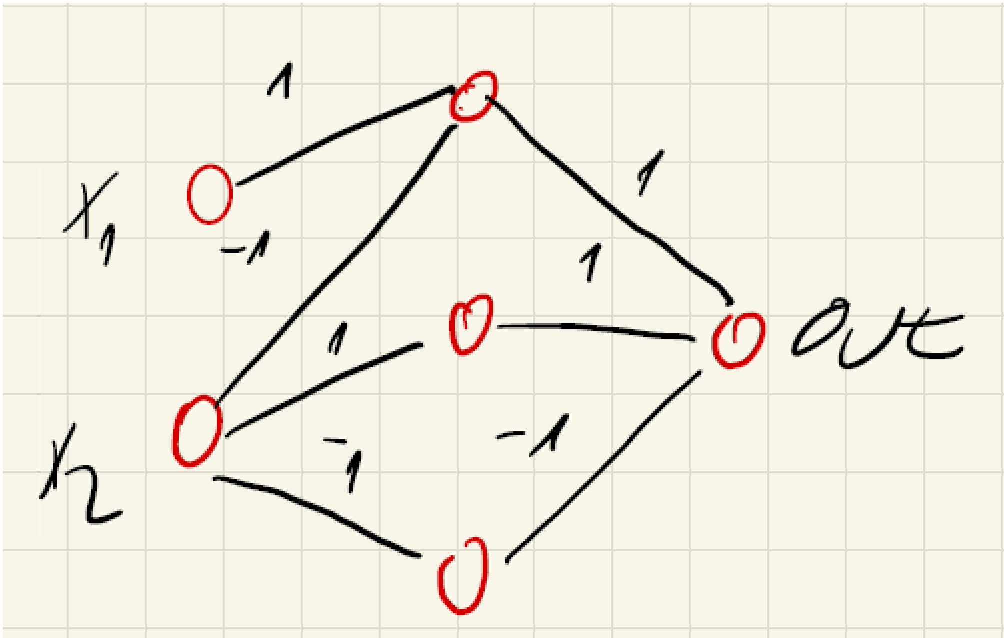
לכן, הצורה של Φ היא:

$$\phi = \begin{pmatrix} 1 & \binom{q}{1}^{0.5} x_1 & \binom{q}{2}^{0.5} x_1^2 & \dots & \binom{q}{q}^{0.5} x_1^q \\ 1 & \binom{q}{1}^{0.5} x_2 & \binom{q}{2}^{0.5} x_2^2 & \dots & \binom{q}{q}^{0.5} x_2^q \\ & & \vdots & & \\ 1 & \binom{q}{1}^{0.5} x_n & \binom{q}{2}^{0.5} x_n^2 & \dots & \binom{q}{q}^{0.5} x_n^q \end{pmatrix}$$

פעולות אלמנטריות אינן משנות את דרגת המטריצה. לכן, אם נחלק כל עמודה j ב $\binom{q}{j}^{0.5}$ נקבל את המטריצה הכתובה בתחילת התרגיל. לכן, קיבלנו שΦ ממםד n.

<u>שאלה 2</u>

נממש את הפונקציה בעזרת שכבה אחת ו3 נוירונים. ניצור את הרשת הבאה:



נקבל שהמוצא יהיה שווה ל$ReLU(x_1 - x_2) + ReLU(x_2) - ReLU(-x_2)$.

נראה ש $x = ReLU(x) - ReLU(-x)$:

אם $x \geq 0$ נקבל ש

$$x \leftarrow \begin{cases} ReLU(x) = x \\ ReLU(-x) = 0 \end{cases}$$

אם $x < 0$ נקבל ש

$$x \leftarrow \begin{cases} ReLU(-x) = x \\ ReLU(x) = 0 \end{cases}$$

לכן, בסה"כ המוצא יהיה ל$ReLU(x_1 - x_2) + x_2$

אם $x_1 > x_2$ אזי $ReLU(x_1 - x_2) = x_1 - x_2$ והמוצא יהיה:

$$x_1 - x_2 + x_2 = x_1$$

אם $x_1 < x_2$ אזי $ReLU(x_1 - x_2) = 0$ והמוצא יהיה:

$$0 + x_2 = x_2$$

בסה"כ קיבלנו $\max\{x_1, x_2\}$.

שאלה 3

סעיף א

נרצה להראות שבעזרת האלגוריתם הנתון נקבל שגיאה של $\frac{1}{4}$ לפחות. נשתמש באלגוריתם הבא:

- נקבל כקלט (S, A).
- אם כל הדוגמאות בS מתויגות בc נחזיר שהוא עלה.
- אם A היא ריקה אזי נחזיר עלה d כאשר d הוא רוב התיוגים בS.
- נגדיר $j = argmax_{i \in A} G(S, i)$ ו$m = majority\ in\ S$.
- נגדיר $S_0 = \{(x, y) \in S \mid x_i = 0\}, S_1 = \{(x, y) \in S \mid x_1 = 0\}$.
- אם אחד מהם קבוצה ריקה, נגדיר את הצד שלו כm. אחרת, נקרא לו בצורה רקורסיבית.
$L_i = (S_0, A \backslash j)$
- נחזיר את העץ הנוצר בצורה הזו.

נאחל $A = \{0,1,2\}$. היא איננה ריקה ויש נקודות בS שמתויגות בצורה שונה(כלומר, לא כולן זהות).
באיטרציה הראשונה נקבל:

$$\begin{aligned} j = \arg\max_{i \in A} G(S, i) = \arg\max_{i \in A} H(Y) - H(Y|X_i) \\ = \arg\max_{i \in A} \mathbb{P}_S[Y = 1] \log(\mathbb{P}_S[Y = 1]) - \mathbb{P}_S[Y = 0] \log(\mathbb{P}_S[Y = 0]) \\ - \mathbb{P}_S[X_i = 1] \mathbb{P}_S[Y = 1|X_i = 1] \log(\mathbb{P}_S[Y = 1|X_i = 1]) \\ - \mathbb{P}_S[X_i = 1] \mathbb{P}_S[Y = 0|X_i = 1] \log(\mathbb{P}_S[Y = 0|X_i = 1]) \\ - \mathbb{P}_S[X_i = 0] \mathbb{P}_S[Y = 1|X_i = 0] \log(\mathbb{P}_S[Y = 1|X_i = 0]) \\ - \mathbb{P}_S[X_i = 0] \mathbb{P}_S[Y = 0|X_i = 0] \log(\mathbb{P}_S[Y = 0|X_i = 0]) \end{aligned}$$

החלק הראשון איננו תלוי ב i ולכן נוכל לחשבו ולקבל:

$$\mathbb{P}_S[Y = 1] = \frac{1}{2}, \mathbb{P}_S[Y = 0] = \frac{1}{2}$$

לכן החלק הראשון יהיה שווה ל

$$\frac{1}{2} \log\left(\frac{1}{2}\right) - \frac{1}{2} \log\left(\frac{1}{2}\right) = 0$$

נשים לב ש:

$$\mathbb{P}_S[X_0 = 0] = \frac{1}{4}, \mathbb{P}[X_1 = 0] = \mathbb{P}[X_2 = 0] = \frac{1}{2}$$

מסימטריה עבור $i = 1,2$ נקבל:

$$\mathbb{P}_S[Y = 1|X_i = 1] = \mathbb{P}_S[Y = 1|X_i = 0] = \mathbb{P}_S[Y = 0|X_i = 1] = \mathbb{P}_S[Y = 0|X_i = 0] = \frac{1}{2}$$

ועבור $i = 0$:

$$\mathbb{P}_S[Y = 1|X_0 = 1] = \frac{2}{3}, \mathbb{P}_S[Y = 1|X_0 = 0] = 0, \mathbb{P}_S[Y = 0|X_0 = 1] = \frac{1}{3}, \mathbb{P}_S[Y = 0|X_0 = 0] = 1$$

עבור $i = 1,2$ נקבל:

$$H(Y)-H(Y|X_i)=-\frac{4}{2}\cdot\frac{1}{2}\log\left(\frac{1}{2}\right)=-\log\left(\frac{1}{2}\right)=0.301$$

עבור $i=0$:

$$H(Y)-H(Y|X_0)=-\frac{3}{4}\cdot\frac{2}{3}\cdot\log\left(\frac{2}{3}\right)-\frac{3}{4}\cdot\frac{1}{3}\log\left(\frac{1}{3}\right)-\frac{1}{4}\cdot 0-\frac{1}{4}\cdot 1\log(1)=-\frac{1}{2}\log\left(\frac{2}{3}\right)-\frac{1}{4}\log\left(\frac{1}{3}\right)$$
$$=0.4774$$

כלומר האלגוריתם ייבחר את x_0. כעת אנו נשארים עם 4 אפשרויות עבור שאר ההחלטות של העץ. לא משנה מה נבחר עבור העלה שיש לו שתי אפשרויות בכל עץ, נקבל שיש נקודה אחת במערך האימונים שמסווג באופן שגוי.

לכן, נקבל עץ עם שגיאה של לפחות $\frac{1}{4}$.

<u>סעיף ב</u>

נתחיל עם מציאת עץ החלטות שמניב שגיאת אימון של 0.אם ניקח את x_0,x_1,x_2 ונשאל קודם על x_2 נקבל:

$$\{((1,1,1),1),((0,0,1),0)\},\{((1,0,0),1),((1,1,0),0)\}$$

נוכל אחר כך לשאול על x_1 ולקבל 4 קבוצות שונות, אחד עבור כל עלה ולקבל שגיאת אימון 0.

שאלה 4

סעיף א

מתכונות ההסתברות של D נקבל:

$$\mathbb{P}_{x\sim D_{t+1}}[h_t(x)\neq y]=\sum_{i=1}^{n}D_{t+1}(i)\cdot\mathbb{1}[h_t(x_i)\neq y_i]=\sum_{i=1}^{n}\frac{D_t(i)e^{-w_t y_i h_t(x_i)}}{Z_t}\cdot\mathbb{1}[h_t(x_i)\neq y_i]$$
$$=\frac{1}{Z_t}\sum_{\forall i:h_t(x_i)\neq y_i}^{n}D_t(i)e^{w_t}=\frac{e^{w_t}}{Z_t}\sum_{\forall i:h_t(x_i)\neq y_i}^{n}D_t(i)=\frac{e^{w_t}}{Z_t}\epsilon_t$$

ראינו ש:

$$Z_t=2\sqrt{\epsilon_t(1-\epsilon_t)},w_t=0.5\log\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$$

לכן נקבל:

$$\mathbb{P}_{x\sim D_{t+1}}[h_t(x)\neq y]=\frac{\sqrt{\frac{1-\epsilon_t}{\epsilon_t}}}{2\sqrt{\epsilon_t(1-\epsilon_t)}}\cdot\epsilon_t=\frac{1}{2\epsilon_t}\cdot\epsilon_t=\frac{1}{2}$$

סעיף ב

נניח בשלילה ש$h_t=h_{t+1}$. יתקיים:

$$\mathbb{P}_{x\sim D_{t+1}}[h_t(x)\neq y]=\mathbb{P}_{x\sim D_{t+1}}[h_{t+1}(x)\neq y]=\epsilon_{t+1}=\frac{1}{2}$$

אבל באלגוריתם אנו יודעים שלכל T קיים $\gamma>0$ כך ש$\epsilon_T<\frac{1}{2}-\gamma<\frac{1}{2}$, בסתירה. לכן$h_t\neq h_{t+1}$.

שאלה 5

סעיף א

ניעזר ברמז $\mathbb{E}\left[y_i \sum_{j=1}^{k} a_j h_j(x_i)\right] \geq \gamma$. מליניאריות התוחלת אנו מקבלים $\sum_{j=1}^{k} a_j \mathbb{E}\left[y_i h_j(x_i)\right] \geq \gamma$. נניח בשלילה שלכל j מתקיים $\Pr_{i\sim D}\left[h_j(x_i) \neq y_i\right] > \frac{1-\gamma}{2}$. לכן:

$$\Pr_{i\sim D}\left[h_j(x_i) = y_i\right] = 1 - \Pr_{i\sim D}\left[h_j(x_i) \neq y_i\right] < \frac{1+\gamma}{2}$$

ונובע:

$$\mathbb{E}\left[y_i h_j(x_i)\right] = -1 \cdot \Pr_{i\sim D}\left[h_j(x_i) \neq y_i\right] + 1 \cdot \Pr_{i\sim D}\left[h_j(x_i) = y_i\right] = \Pr_{i\sim D}\left[h_j(x_i) = y_i\right] - \Pr_{i\sim D}\left[h_j(x_i) \neq y_i\right] < \frac{1+\gamma}{2} - \frac{1-\gamma}{2} = \gamma$$

ונקבל:

$$\mathbb{E}\left[y_i \sum_{j=1}^{k} a_j h_j(x_i)\right] = \sum_{j=1}^{k} a_j \mathbb{E}\left[y_i h_j(x_i)\right] < \sum_{j=1}^{k} a_j \gamma = \gamma$$

בסתירה.

סעיף ב

נגדיר $k = 4d - 1$ ונבחר את ההיפותזות שלנו $h_1, \dots, h_{4d-1}$. $2d - 1$ יהיו קבועות ויחזירו 1- עבור כל קלט. כל השאר יסמנו את הממד הd של מצולע המשוכלל שלנו שאותו אנו נאמן לשגיאה 0. נבחר $\forall 1 \leq j \leq 4d-1: a_j = \frac{1}{4d-1}$.

מתקיים:

$$y_i \sum_{j=1}^{k} a_j h_j(x_i) = y_i \sum_{j=1}^{4d-1} \frac{1}{4d-1} h_j(x_i) = y_i \left(-\frac{2d-1}{4d-1} + \sum_{j=2d}^{4d-1} \frac{h_j(x_i)}{4d-1}\right)$$

אם $y_i = 1$ אז $\forall 2d \leq j \leq 4d-1: h_j(x_i) = 1$ ונקבל:

$$y_i \sum_{j=1}^{k} a_j h_j(x_i) = 1\left(-\frac{2d-1}{4d-1} + \frac{2d}{4d-1}\right) = \frac{1}{4d-1}$$

אם $y_i = -1$ אז הקלט איננו על המצולע מה שאומר שלפחות אחד הקואורדינטות מחוץ למצולע. כלומר

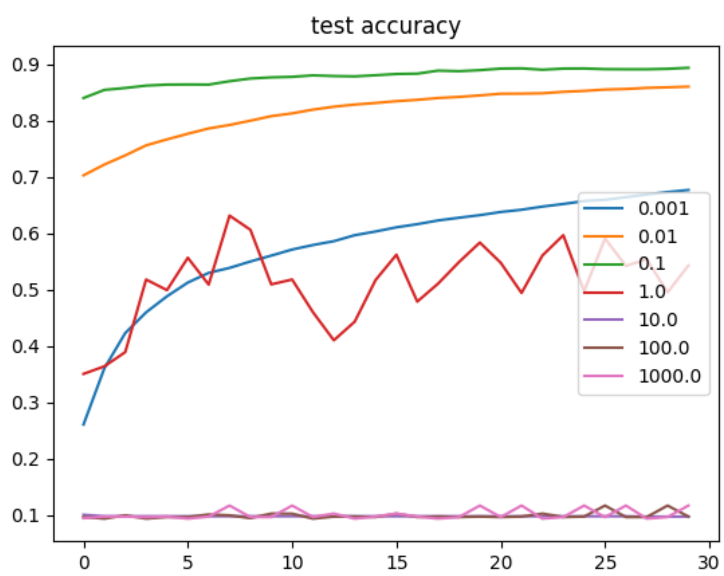
$$\exists 2d \leq j \leq 4d-1: h_j(x_i) = -1$$

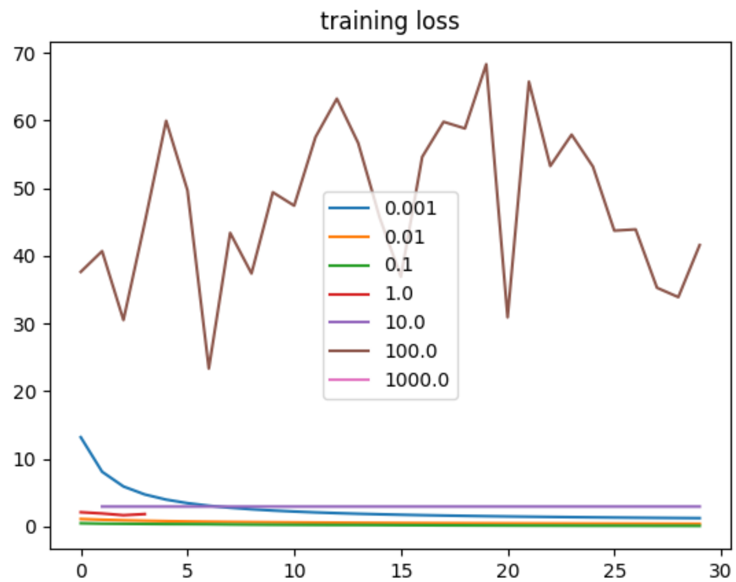
$$y_i \sum_{j=1}^{k} a_j h_j(x_i) \geq -1\left(-\frac{2d-1}{4d-1} + \frac{2d-1}{4d-1} - \frac{1}{4d-1}\right) = \frac{1}{4d-1}$$

נבחר $\gamma = \frac{1}{4d-1}$ ונקבל את מה שרצינו.

<u>חלק תכנותי</u>

<u>סעיף ב</u>





ניתן לראות שאנו מקבלים את התוצאות הטובות ביותר עבור קצב של 0.1. אם נעשה צעד שהוא קטן מידי אז העלייה תהיה "איטית" מידי ולא נתכנס לערך אופטימלי במהלך האיטרציות שלנו. מצד שני, אם נעשה צעדים גדולים מידי אנחנו נהיה בover-fitting ולא נצליח להתכנס לערך האופטימלי גם כן.

<u>סעיף ג</u>

הגענו לדיוק של יותר מ93 אחוז.

```
Epoch 14 test accuracy: 0.9304
Epoch 15 test accuracy: 0.9315
Epoch 16 test accuracy: 0.9321
Epoch 17 test accuracy: 0.9315
Epoch 18 test accuracy: 0.9316
Epoch 19 test accuracy: 0.9342
Epoch 20 test accuracy: 0.9352
Epoch 21 test accuracy: 0.9349
Epoch 22 test accuracy: 0.9353
Epoch 23 test accuracy: 0.9363
Epoch 24 test accuracy: 0.9356
Epoch 25 test accuracy: 0.9361
Epoch 26 test accuracy: 0.936
Epoch 27 test accuracy: 0.936
Epoch 28 test accuracy: 0.9369
Epoch 29 test accuracy: 0.9359
```