מבוא ללמידה חישובית – תרגיל 5

.1

לפי הלמה מהתרגול אם המטריצה:

$$X = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix}$$

מקיימת ש $rank(X) = n$ אז הדאטה ניתן להפרדה לינארית ואם הדאטה ניתן להפרדה לינארית אז hard svm משיג על דאטה האימון שגיאה אפס.

לכן מספיק להוכיח כי עבור הגרעין $K_q(x, x') = (1 + xx')^q$ מתקיים שהדרגה של המטריצה

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

היא n

מתקיים :

$$K_q(x, x') = \sum_{j=1}^{Q} \binom{q}{j} (xx')^j = \sum_{j=1}^{Q} \binom{q}{j}^{\frac{1}{2}} x^j \binom{q}{j}^{\frac{1}{2}} x'^j \Rightarrow$$

$$\phi(x) = \begin{pmatrix} \binom{q}{0} x^0 \\ \binom{q}{1}^{\frac{1}{2}} x \\ \vdots \\ \binom{q}{q}^{\frac{1}{2}} x^q \end{pmatrix}$$

$$rank \begin{pmatrix} x_1^0 & x_1 & \dots & x_1^q \\ x_2^0 & x_2 & \dots & x_2^q \\ \vdots & \vdots & & \vdots \\ x_n^0 & x_n & \dots & x_n^q \end{pmatrix} = n$$

כי דרגה של מטריצת ונדרומונד היא n , עבור כל j נכפיל כל עמודה בקבוע $\binom{q}{j-1}^{\frac{1}{2}}$ הדרגה לא תשתנה ונקבל את המטריצה

$$\begin{pmatrix} x_1^0 & \binom{q}{1}^{\frac{1}{2}} x_1 & \cdots & \binom{q}{q}^{\frac{1}{2}} x_1^q \\ x_2^0 & \binom{q}{1}^{\frac{1}{2}} x_2 & \cdots & \binom{q}{q}^{\frac{1}{2}} x_2^q \\ \vdots & \vdots & & \vdots \\ x_n^0 & \binom{q}{1}^{\frac{1}{2}} x_n & \cdots & \binom{q}{q}^{\frac{1}{2}} x_n^q \end{pmatrix} = \Phi$$

ניתן להפריד את הדאטה על ידי מפריד לינארי ו*hard svm* מקבל טעות אפס על הדאטה.

.2

מתקיים :

$$\max\{x_1, x_2\} = \frac{1}{2}(x_1 + x_2) + \frac{1}{2}|x_1 - x_2|$$

נבנה רשת בעלת שכבה חבויה אחת כאשר פונקצית האקטיבציה היא *ReLU*

$$W_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ -1 & -1 \\ -1 & 1 \end{pmatrix}, W_2 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

$$z_0 = \boldsymbol{x} = \binom{x_1}{x_2} \to z_1 = \sigma(W_1\boldsymbol{x}) = \begin{pmatrix} \max\{x_1 + x_2, 0\} \\ \max\{x_1 - x_2, 0\} \\ \max\{x_2 - x_1, 0\} \\ \max\{-x_1 - x_2, 0\} \end{pmatrix} \to output = \sigma(W_2 z_1)$$

$$= \frac{1}{2}\max\{x_1 + x_2, 0\} + \frac{1}{2}\max\{x_1 - x_2, 0\} + \frac{1}{2}\max\{x_2 - x_1, 0\} - \frac{1}{2}\max\{-x_1 - x_2, 0\}$$

נשים לב שאם מתקיים $x_1 + x_2 > 0 \; or - (x_1 + x_2) > 0 \; and \; x_1 - x_2 > 0 \; or \; x_2 - x_1 > 0$ לכן:

$$= \frac{1}{2}(x_1 + x_2) + \frac{1}{2}|x_1 - x_2| = \max\{x_1, x_2\}$$

.3

(a) נראה שהשגיאה האמפירית היא לפחות $\frac{1}{4}$.

נמצא את עץ החלטה על ידי האלגוריתם עבור קבוצת האימון S ו$A = [3]$, ופונקציית הפוטנציאל

$$val(q) = -\frac{1}{2}[q\log q + (1-q)\log(1-q)]$$

באיטרציה הראשונה נבחר $j = \arg\max_{i \in A} Gain(S, i) = 1$ כי:

$$Gain(S,1) = val\left(\frac{1}{2}\right) - \frac{3}{4}val\left(\frac{2}{3}\right) - \frac{1}{4}val\left(\frac{1}{2}\right) = \frac{1}{2} - \frac{3}{4}(-0.276) - \frac{1}{4}\left(\frac{1}{2}\right) = 0.582$$

$$Gain(S,2) = val\left(\frac{1}{2}\right) - \frac{1}{2} val\left(\frac{1}{2}\right) - \frac{1}{2} val\left(\frac{1}{2}\right) = 0$$
$$Gain(S,3) = val\left(\frac{1}{2}\right) - \frac{1}{2} val\left(\frac{1}{2}\right) - \frac{1}{2} val\left(\frac{1}{2}\right) = 0$$

כעת נבצע את האלגוריתם עבור 2 תתי הקבוצות לפי האינדקס הראשון:

$$S_0 = \{((0,0,1),0)\}, S_1 = \{((1,1,1),1), ((1,0,0),1), ((1,1,0),0)\}$$

עבור S_0 יש רק את הלייבל 0 לכן זהו עלה.

עבור S_1 j יבחר שרירותית כי:

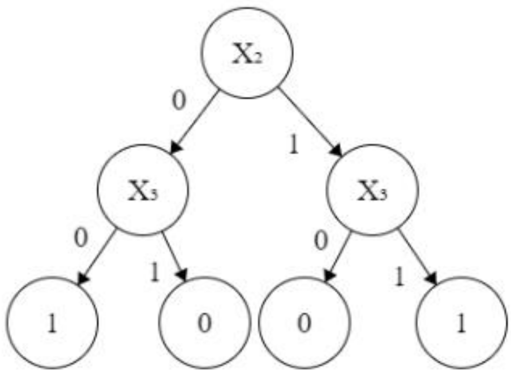
$$Gain(S,2) = val\left(\frac{2}{3}\right) - \frac{2}{3} val\left(\frac{1}{2}\right) - \frac{1}{3} val(1) = -0.276 - \frac{2}{3} - 0 = -0.942$$
$$Gain(S,3) = val\left(\frac{2}{3}\right) - \frac{1}{3} val(1) - \frac{2}{3} val\left(\frac{1}{2}\right) = -0.276 - \frac{2}{3} = -0.942$$

נשים שעבור כל אחת מהאופציות, יהיה איבר עבורו התיוג שגוי :

$$X_2 \to ((1,1,1),1)\ and\ ((1,1,0),0)$$
$$X_3 \to ((1,0,0),1)\ and\ ((1,1,0),0)$$

ובגלל שע"פ השאלה לא ממשיכים לבצע עוד איטרציות, אכן נטעה על לפחות אחד, והשגיאה אכן תיהיה לפחות $\frac{1}{4}$, כדרוש.

(b) העץ בעל שגיאה אפס הוא



4. (a

$$\Pr_{x\sim D_{t+1}}[h_t(x) \neq y] = \sum_{i\in\{i|h_t(x_i) \neq y_i\}} D_{t+1}(i) = \sum_{i\in\{i|h_t(x_i) \neq y_i\}} \frac{D_t(i)e^{-y_i w_t h_t(x_i)}}{Z_t}$$

מכיוון ש $h_t(x) \neq y$ אז $y_i h(x_i) = -1$ לכן:

$$\sum_{i\in\{i|h_t(x_i)\neq y_i\}} \frac{D_t(i)e^{-y_i w_t h_t(x_i)}}{Z_t} = \sum_{i\in\{i|h_t(x_i)\neq y_i\}} \frac{D_t(i)e^{w_t}}{Z_t} = \frac{e^{w_t}}{Z_t} \sum_{i\in\{i|h_t(x_i)\neq y_i\}} D_t(i)$$

$$= \frac{e^{w_t}}{Z_t}\epsilon_t = \sqrt{\frac{1-\epsilon_t}{\epsilon_t}}\frac{\epsilon_t}{Z_t} = \sqrt{\frac{\epsilon_t(1-\epsilon_t)}{2\epsilon_t(1-\epsilon_t)}} = \frac{1}{2}$$

כאשר שני המעברים האחרונים נובעים מהגדרה של w_t ומלמה 9.3.4

b) ראינו בכיתה כי $h_{t+1} = WL(D_{t+1}, S)$ נניח בשלילה שמתקיים $h_t = h_{t+1}$ מסעיף א

$$\Pr_{x\sim D_{t+1}}[h_t(x)\neq y] = \frac{1}{2} \Rightarrow e_{t+1} = \frac{1}{2}$$

סתירה לכך שh_{t+1} הוא לומד חלש. לכן $h_t \neq h_{t+1}$

5. *a*)

$$\gamma \leq y_i \sum_{j=1}^{k} a_j h_j(x_i) \Rightarrow \mathbb{E}[\gamma] \leq \mathbb{E}\left[y_i \sum_{j=1}^{k} a_j h_j(x_i)\right] = \sum_{i=1}^{n} D(i) y_i \sum_{j=1}^{k} a_j h_j(x_i)$$

$$= \sum_{j=1}^{k} a_j \sum_{i=1}^{n} D(i) y_i\, h_j(x_i)$$

נגדיר

$$j^* = \arg\max_{j} \sum_{i=1}^{n} D(i) y_i h_j(x_i)$$

נניח בשלילה

$$\sum_{i=1}^{n} D(i) y_i h_{j^*}(x_i) < \gamma$$

כלומר

$$\sum_{j=1}^{k} a_j \sum_{i=1}^{n} D(i) y_i h_j(x_i) < \sum_{j=1}^{k} a_j \sum_{i=1}^{n} D(i) y_i h_{j^*}(x_i) = \left(\sum_{i=1}^{n} D(i) y_i h_{j^*}(x_i)\right) \sum_{j=1}^{k} a_j$$

$$= \sum_{i=1}^{n} D(i) y_i h_{j^*}(x_i) < \gamma$$

סתירה. כלומר קיים j המקיים

$$\sum_{i=1}^{n} D(i) y_i h_j(x_i) \geq \gamma$$

אז מתקיים

$$
\begin{aligned}
\sum_{i=1}^{n} D(i) y_i h_j(x_i) &= \sum_{i:y_i=h_t(x_i)} D(i) y_i h_j(x_i) + \sum_{i:y_i\neq h_t(x_i)} D(i) y_i h_j(x_i) \\
&= \sum_{i:y_i=h_t(x_i)} D(i) - \sum_{i:y_i\neq h_t(x_i)} D(i) \\
&= \left(1 - \mathbb{P}_{i\sim D}\left[h_j(x_i) \neq y_i\right]\right) - \mathbb{P}_{i\sim D}\left[h_j(x_i) \neq y_i\right] \\
&= 1 - 2\mathbb{P}_{i\sim D}\left[h_j(x_i) \neq y_i\right] \geq \gamma \Rightarrow \mathbb{P}_{i\sim D}\left[h_j(x_i) \neq y_i\right] \leq \frac{1}{2} - \frac{\gamma}{2}
\end{aligned}
$$

b) נגדיר $\forall i\ a_i = \frac{1}{4d-1}$ ו $k = 4d-1$ נשים לב שמתקיים $\sum_{i=1}^{k} a_i = 1$

תהי d-היפר תיבה $B = [b_1, c_1] \times \ldots \times [b_d, c_d]$ שמסווגת את הנקודות נכון

נגדיר $h_{b_1}, h_{c_1}, h_{b_2}, h_{c_2}, \ldots, h_{b_d}, h_{c_d}, h_{2d+1}, \ldots, h_{4d-1}$

כאשר:

$$
h_{b_j}(x) = \begin{cases} 1 & x_j \geq b_j \\ -1 & x_j < b_j \end{cases}
$$

$$
h_{c_j}(x) = \begin{cases} -1 & x_j > c_j \\ 1 & x_j \leq c_j \end{cases}
$$

$$
h_j = -1 \quad k \geq 2k+1
$$

נשים לב אם $x_i \in B$ אז $y_i = 1$

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

אם $x \notin B$ אז $y_i = -1$

נשים לב ש $y_i \sum_{j=1}^{k} a_j h_j(x_i)$ מקסימלי כאשר רק היפותזה אחרת מתוך $h_{b_1}, h_{c_1}, \ldots, h_{b_d}, h_{c_d}$

מחזירה *1*- לכן

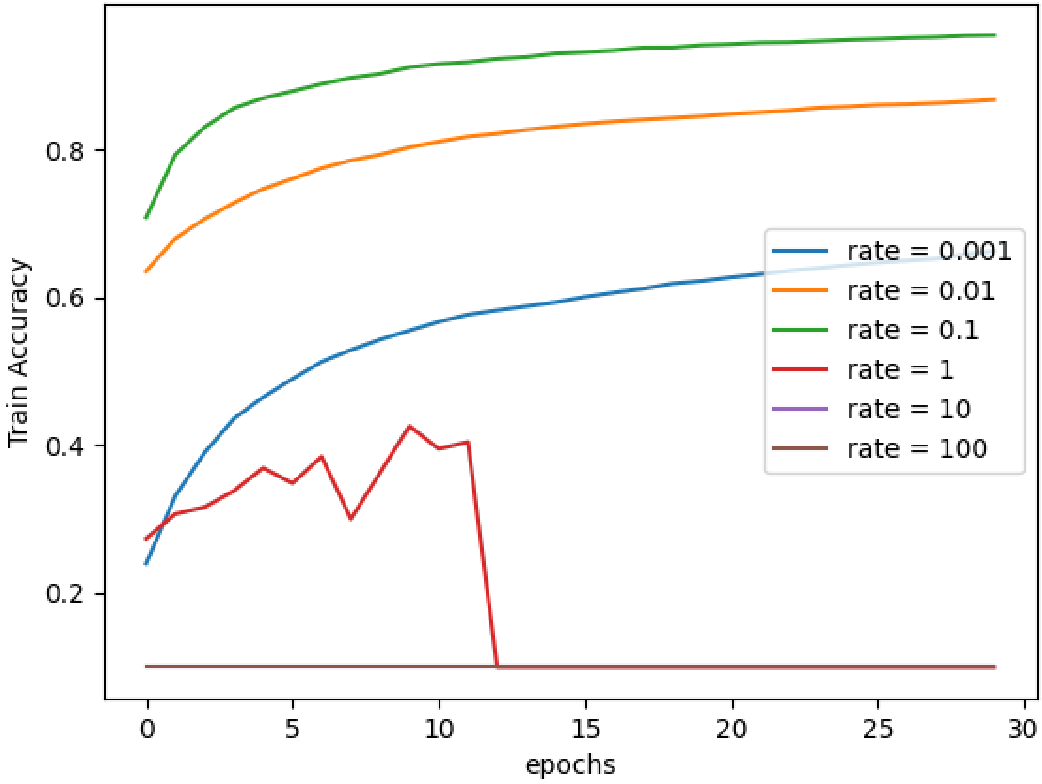
$$
\begin{aligned}
y_i \sum_{j=1}^{k} a_j h_j(x_i) &\leq \frac{1}{4d-1}\left(\sum_{i:y_i=h_t(x_i)} 1 + \sum_{i:y_i\neq h_t(x_i)} -1\right) = \frac{1}{4d-1}(2d - (2d-1) \\
&= \frac{1}{4d-1}
\end{aligned}
$$

נגדיר $\gamma = \frac{1}{4d-1}$ ונקבל לכל x

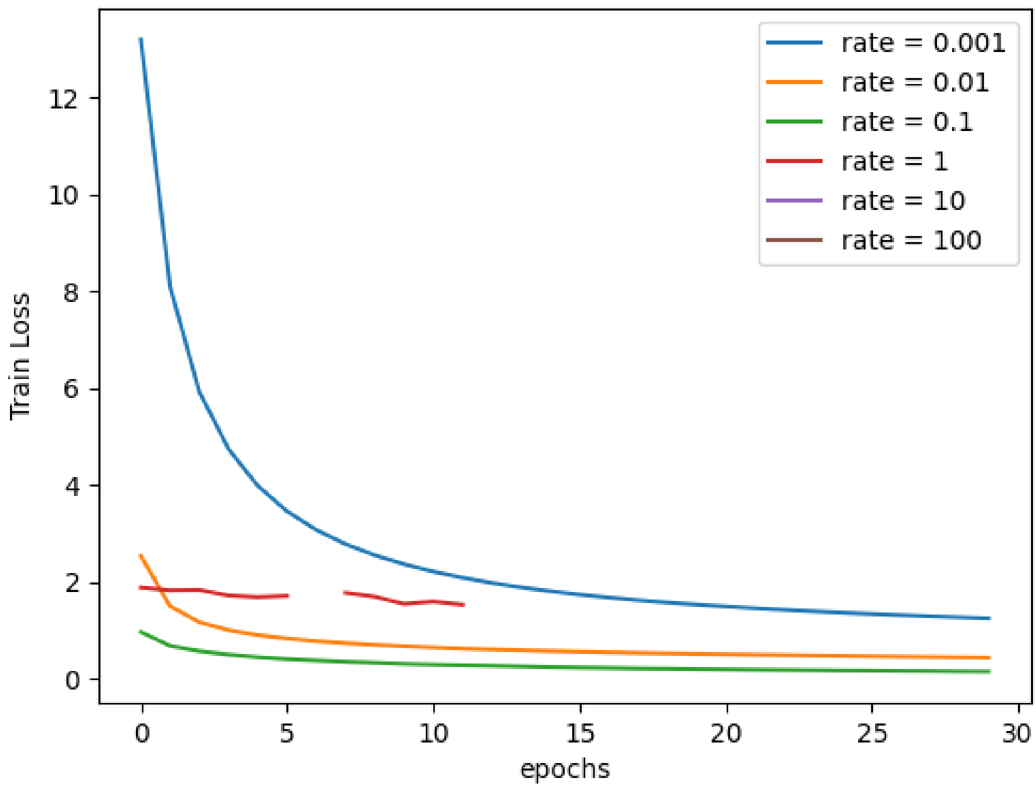
$$
y \sum_{j=1}^{k} a_j h_j(x) \leq \gamma
$$

החלק המעשי

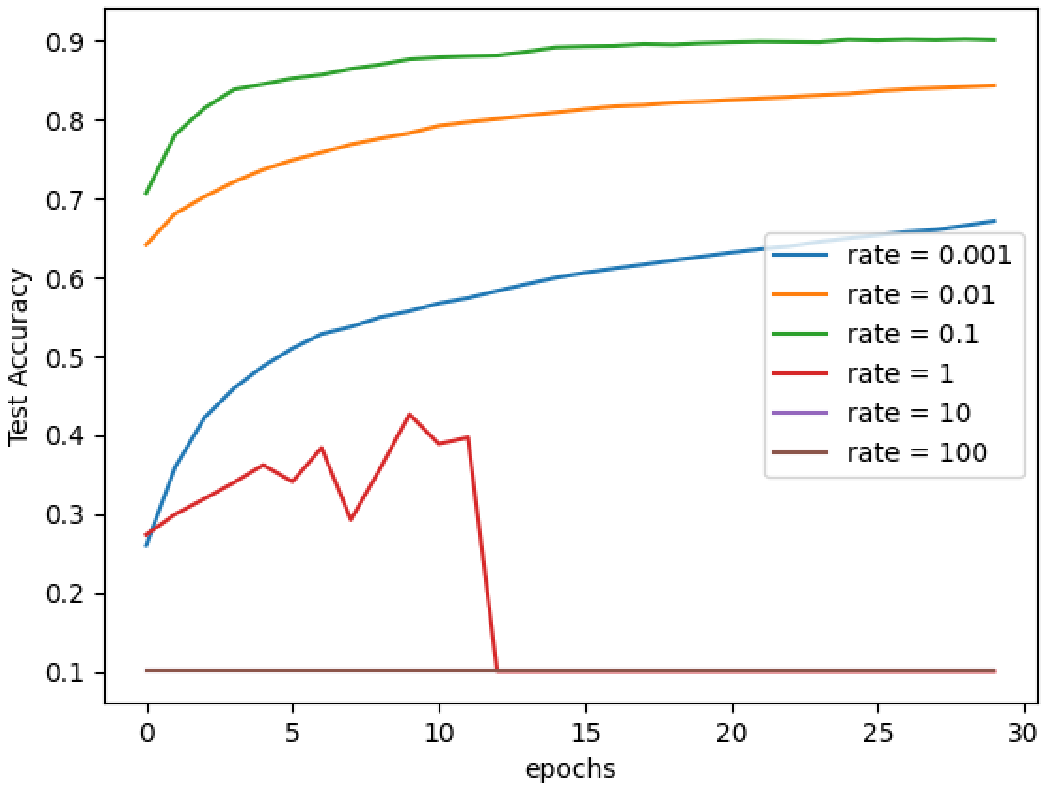
.2



0.8
0.6
0.4
0.2
Train Accuracy
0
5
10
15
20
25
30
epochs
rate = 0.001
rate = 0.01
rate = 0.1
rate = 1
rate = 10
rate = 100



rate = 0.001
rate = 0.01
rate = 0.1
rate = 1
rate = 10
rate = 100
Train Loss
epochs
0
2
4
6
8
10
12
0
5
10
15
20
25
30



rate = 0.001
rate = 0.01
rate = 0.1
rate = 1
rate = 10
rate = 100
Test Accuracy
epochs
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
0
5
10
15
20
25
30

ניתן לראות כאשר קצב המידה קטן מידי הרשת לא מצליחה להגיע לביצועים אופטימלים ניתן להבין שגודל הצעד קטן מידי הרשת לא מגיעה ל"עמק" הנמוך ביותר מנגד כאשר קצב הלמידה גובהה הביצועים גרועים זאת מכיוון שהאולי "מדלגים" על המינימום ולא מצלחים להגיע קרוב מספיק

.3

הדיוק המתקבל הוא 0.93