דוח הכנה 2 – 023 מערכות לומדות

נתנאל רוטשילד 204937841
גיא פרג 201096534

שאלה 1)

Input$(X_{TRAIN}, Y_{TRAIN}, x_{sample})$

Set $\min - distance = \infty$

Set $class = null$

For $x_{train} \in X_{TRAIN}$

If $|x_{train} - x_{sample}| < min - dstance$

Set $\min - distance = |x_{train} - x_{sample}|$

Set $class = x_{train}.class$

Return class

סיבוכיות זמן הריצה עבור הקוד שלנו הוא כגודל הקלט $O(n)$. נדרש לעבור על כל אחת מהדגימות למצוא את המרחק המינימלי מבניהם.

שאלה 2) ניתן למיין את סדרת הלמידה מראש כך שמציאת מקומו של הדגימה הבודדת תהיה בסיבוכיות $O(\log n)$

Sort_Func(X_{TRAIN}, Y_{TRAIN})

Set $Array[0][0] = x_{train_1}, y_{train_1}$

Set $X_{TRAIN} = x_{train_1} \notin X_{TRAIN}$

For $i \ in \ sizeof(X_{TRAIN})$

Find $x_i = min - distance(x_{train_1}, X_{TRAIN})$

Set $Array[i][i] = x_i, y_i$

Set $X_{TRAIN} = x_i \notin X_{TRAIN}$

Return $Array$

Input($Array_{TRAIN}, x_{sample}$)

$DistanceFun = \ |x - x`|$

Set $class =$ Find_Shortest_Distance($Array_{TRAIN}, x_{sample}, DistanceFunc$)

Return class

כאשר המיון הראשני נעשה בסיבובוכיות $n \cdot log\, n$ והסיווג עבור נקודה חדשה נעשה בסיבוכיות $\log n$

שאלה 3) נפריח את הטענה כי המרחק הזוויתי מקיים את הגדרת פונקציית המרחק בכל ממד $d > 1$.

נבחן עבור $d = 2$ נמצא 3 נקודות כך שיהיה סתירה בניסיון למיין אותם לפי מרחק.

$$a = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, b \ = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, c = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}$$

כעת נחשב את המרחקים בין כל זוג נקודות:

$$d(a,b) = 1 - \frac{(0\ 1)^T \cdot (1\ 0)}{1} = 1$$

$$d(a,c) = 1 - \frac{(0\ 1)^T \cdot \left(\frac{\sqrt{2}}{2}\ \frac{\sqrt{2}}{2}\right)}{1} = 1 - \frac{\sqrt{2}}{2}$$

$$d(c,b) = 1 - \frac{(2\ 0)^T \cdot \left(\frac{\sqrt{2}}{2}\ \frac{\sqrt{2}}{2}\right)}{1} = 1 - \frac{\sqrt{2}}{2}$$

הגענו לסתירה של מיון לפי מרחק:

$$d(a,c) + d(c,b) < d(a,b)$$

כלומר המעבר בין $a \rightarrow c \rightarrow b$ קצר יותר מ: $a \rightarrow b$ ולכן אם ננסה למיין לפי מרחק נתקל בבעיה כי המרחק שנתון על ידי הנוסחה לא תיתן לנו את המרחק המינימלי.

שאלה 4) הפרמטר אלפא קובע לנו את "גודל" הצעד שנעשה כדי לקרב את המישור שלנו למיקומו הנכון ביחס לסט למידה. במקרה ונבחר בצעד "קטן מידי" עלול להיות מצב בו מעבר יחיד על סדרת האימון לא תספיק כדי לתת לנו מישור מתאים, לכן נדרוש מעבר מספר פעמים.

5. נניח שדוגמת אימון (x,y) בעלת תיוג $y=+1$ מסווגות באופן שגוי ידי פרספטרון עם וקטור משקלים w^t, כלומר $\hat{y}=\varphi_{HL}\left(w^{tT}x\right)=-1$. הראו כי עבור $\alpha>0$ גדול מספיק, וקטור המשקלים המעודכן w^{t+1} יוביל לסיווג נכון, כלומר $\hat{y}=\varphi_{HL}\left(w^{t+1T}x\right)=+1$.

נתונה דוגמת אימון (x,y) בעלת תיוג $y=1$ המסווגת באופן שגוי עבור וקטור המשקלים w^t, כלומר:

$$\hat{y}=\varphi_{HL}(w^{tT}x)=-1$$

כמו שראינו בחומר ההכנה מתקיים:

$$y=\varphi_{HL}\left(w^Tx\right)=\begin{cases}-1: & w^Tx<0\\ +1: & w^Tx\geq 0\end{cases}$$

ולכן עבור הנתון

$$\hat{y}=\varphi_{HL}(w^{tT}x)=-1$$

מתקיים כי:

$$w^{tT}x<0$$

לפי שלב 3 באלגוריתם הלמידה – עדכון וקטור המשקלים w^{t+1}:

$$w^{t+1}=\begin{cases}w^t+\alpha\cdot(y-\hat{y})\cdot x & ,y\neq\hat{y}\\ w^t & ,y=\hat{y}\end{cases}$$

עבור דוגמת האימון שלנו:

$$w^{t+1}=w^t+\alpha\cdot(y-\hat{y})\cdot x=w^t+\alpha\cdot\big(1-(-1)\big)\cdot x=w^t+2\alpha x$$

$$\Longrightarrow\ w^{t+1}=w^t+2\alpha x$$

ובשלב הבא:

$$\hat{y}=\varphi_{HL}((w^{t+1})^Tx)\overset{w^{t+1}=w^t+2\alpha x}{\cong}\varphi_{HL}((w^t+2\alpha x)^Tx)=\varphi_{HL}(w^{tT}x+2\alpha xx)=\varphi_{HL}(w^{tT}x+2\alpha x^2)$$

נדרוש כי התקיים:

$$(w^{tT}x+2\alpha x^2)\geq 0$$

$$2\alpha x^2\geq -w^{tT}x$$

$$\alpha\geq-\frac{w^{tT}x}{2x^2}=(-1)\frac{w^{tT}}{2x}$$

בפרט מתקיים עבור

$$\alpha=(-1)\frac{w^{tT}}{2x}$$

$$\hat{y}=\varphi_{HL}((w^{t+1})^Tx)=\varphi_{HL}(w^{tT}x+2\alpha x^2)=\varphi_{HL}\left(w^{tT}x+2\left[(-1)\frac{w^{tT}}{2x}\right]x^2\right)=\varphi_{HL}(0)=1$$

ובפרט עבור $\alpha\geq 0$

6. הראו כי עבור האלגוריתם של רוזנבלט, בחירה של משקלים התחלתיים w^0 וגודל צעד α שקולה לבחירת משקלים התחלתיים w^0/α וגודל צעד 1. במינוח "שקולה" הכוונה לכך ששני המסווגים יתנו תוצאה זהה עבור כל סדרת בוחן.

קראנו בהכנה כי אנו עושים שימוש בפונקציה:

פונקציית ההפעלה φ שבה נבחר בניסוי זה היא פונקציית הפעלה מסוג Hard Limiter, והיא מותאמת *לסיווג בינארי* בין שתי מחלקות:

$$y=\varphi_{HL}\left(w^T x\right)=\begin{cases}-1: & w^T x<0\\ +1: & w^T x\geq 0\end{cases}$$

וראינו בהכנה כי מתקיים $\alpha > 0$ שכן מייצג את גודל הצעד/התקדמות.

ולכן עבור $\alpha > 0$ מתקיים שגם $\alpha^{-1} > 0$

עבור המסווגים שלנו השוני היינו בהכפלה בקבוע (גדול מ-0) – עבור הצעד הראשון נקבל תוצאה זהה.

עבור הצעד הבא

$$w_a^1 = w_a^0 + 2\alpha x$$

$$w_b^1 = \frac{w_a^0}{\alpha} + 2x = (w_a^0 + 2\alpha x)\alpha^{-1} = \frac{w_a^1}{\alpha}$$

עבור הצעד הבא

$$w_a^2 = w_a^1 + 2\alpha x$$

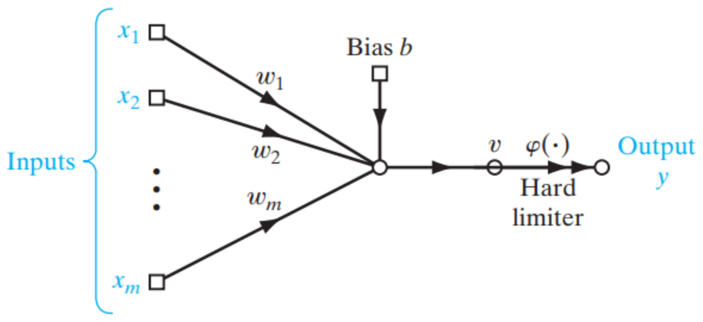
$$w_b^2 = \frac{w_a^1}{\alpha} + 2x = (w_a^1 + 2\alpha x)\alpha^{-1} = \frac{w_a^2}{\alpha}$$

עבור המקרה הכללי (בהנחה כי $y \neq \hat{y}$)

$$w_a^{t+1} = w_a^t + 2\alpha x$$

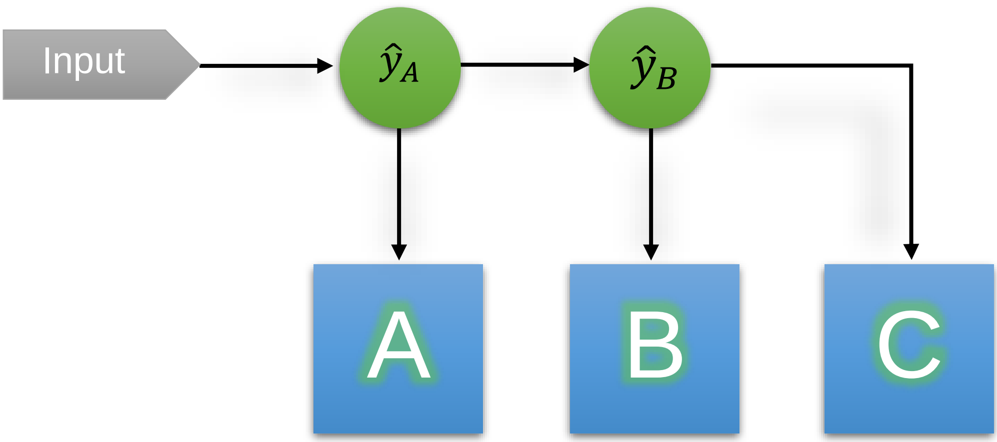
$$w_b^{t+1} = \frac{w_a^t}{\alpha} + 2x = (w_a^t + 2\alpha x)\alpha^{-1} = \frac{w_a^{t+1}}{\alpha}$$

ולכן עבור $\alpha > 0$ ועבור פונקציה $\varphi_{HL}(w^t)$ שהיא *hard_limiter* נקבל עבור על צעד בסדרת האימון והבוחן שוויון. (שכן הסימן לא משתנה)

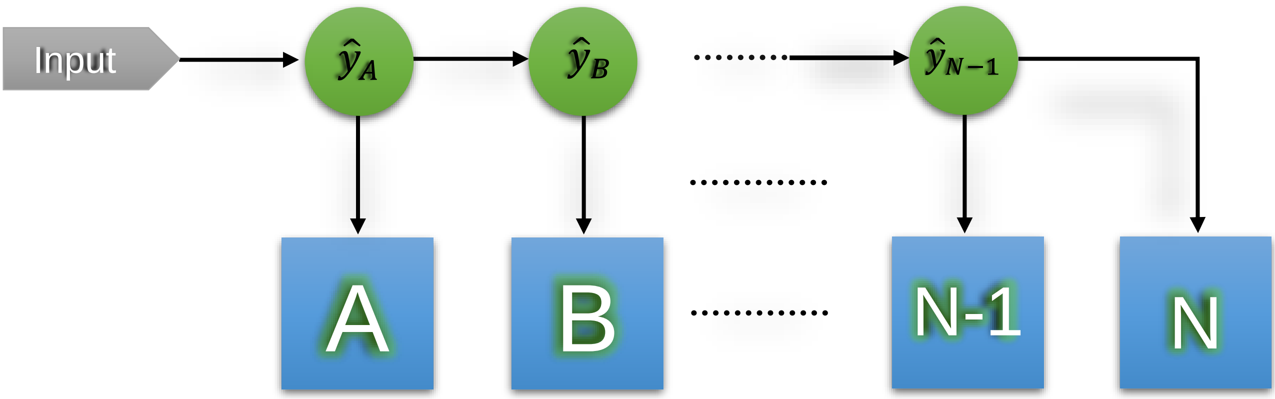


7. הציעו מימוש למסווג תלת מחלקתי הבנוי ממספר מסווגי פרספטרון בינאריים. לכמה מסווגים בינאריים תידרשו על מנת לממש מסווג n-מחלקתי?

- **מסווג תלת מחלקתי** יהיה בנוי משני מסווגי פרספטרון בינאריים, לדוגמא:

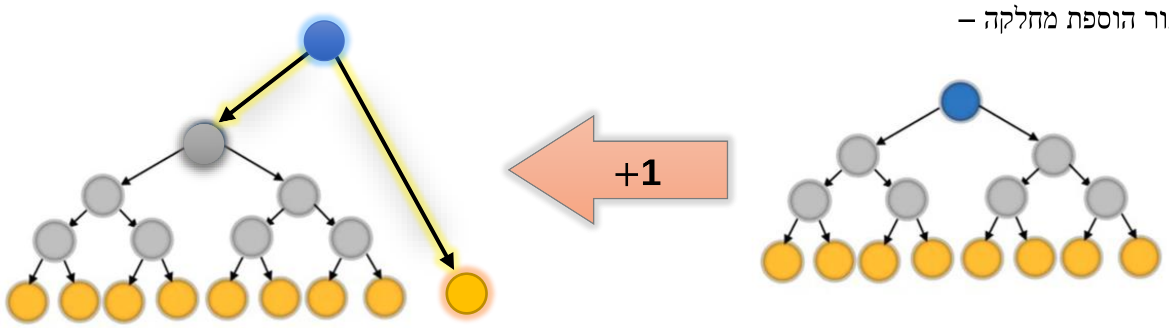


- **עבור המקרה הכללי** -



נסתכל על הבעיה כבנייה של עץ בינארי כאשר:
- ❖ כל עלה של העץ מייצג מחלקה – כלומר נבנה עץ בעל n עלים.
- ❖ כל צומת שהיא לפחות בעלת בן אחד (כלומר לא עלה) נספור כמסווג בינארי.
- ❖ שורש העץ יהווה המסווג אשר יקבל את וקטור הכניסה.

נקבל עבור n עלים (n מחלקות) – עץ בעל $(n-1)$ צמתים פנימיים. (טענה שהוכחה בקורס מבני נתונים)
נשים לב כי עבור הוספת מחלקה נוספת נצטרך להוסיף עוד מסווג אחד כלומר:
עבור הוספת מחלקה –



מסווג החדש יקבע שייכות למחלקות העץ הקודם או למחלקה החדשה שהוספנו.

ולכן למימוש מסווג n-מחלקתי נזדקק ל- $(n-1)$ מסווגי פרספטרון בינאריים