Machine Learning- HW5

Dor Haboosha - 208663534

Itay Golan – 206480402

1. Kernels and mapping functions (30 pts)
   a. (10 pts) Consider two kernels K_1 and K_2, with the mappings φ_1 and φ_2 respectively. Show that $K = 5K_1 + 4K_2$ is also a kernel and find its corresponding φ.
   b. (10 pts) Consider a kernel K_1 and its corresponding mapping φ_1 that maps from the lower space R^n to a higher space R^m (m > n). We know that the data in the higher space R^m, is separable by a linear classifier with the weights vector w.
   Given a different kernel K_2 and its corresponding mapping φ_2, we create a kernel $K = 5K_1 + 4K_2$ as in section a above. Can you find a linear classifier in the higher space to which φ, the mapping corresponding to the kernel K, is mapping?
   If YES, find the linear classifier weight vector.
   If NO, prove why not.
   c. (10 pts) Consider the space $S = \{1,2, \dots N\}$ for some finite N (each instance in the space is a 1-dimension vector and the possible values are 1, 2, …, N) and the function $K(x,y) = 9 \cdot f(x,y)$ for every $x, y \in S$.
   Prove that K is a valid kernel by finding a mapping φ such that:
   $$\varphi(x) \cdot \varphi(y) = 9\min(x,y) = K(x,y)$$
   For example, if the instances are $x = 4, y = 8$, for some $N \geq 8$, then:
   $$\varphi(x) \cdot \varphi(y) = \varphi(4) \cdot \varphi(8) = 9 \cdot \min(4,8) = 36$$

תשובה:

1. a) מאחר ונתון כי K_1 קרנל אזי קיים מיפוי φ_1 שהוא המיפוי של K_1 כך ש-
2. $K_1(x,y) = \varphi_1(x) * \varphi_1(y)$. בנוסף, מאחר ונתון כי K_2 קרנל אזי קיים מיפוי φ_2 שהוא המיפוי של K_2 כך ש-

$K_2(x,y) = \varphi_2(x) * \varphi_2(y)$.

נמצא את המיפוי φ_3 כך ש-

$$K(x,y) = \varphi_3(x) * \varphi_3(y)^t = 5K_1 + 4K_2 = 5 * \left(\varphi_1(x) * \varphi_1(y)\right) + 4(\varphi_2(x) * \varphi_2(y))$$

. נשים לב, כי אנו רוצים בעצם ש-

$\varphi_3(x) * \varphi_3(y)^t = 5 * (\varphi_1(x) * \varphi_1(y)) + 4(\varphi_2(x) * \varphi_2(y))$ לכן נסמן ש-

$\varphi_3(x) = (a * \varphi_1(x), b * \varphi_2(x))$ ונראה כי

$$\begin{aligned}\varphi_3(x) * \varphi_3(y) &= \left(a\varphi_1(x), b\varphi_2(y)\right) * \left(a\varphi_1(x), b\varphi_2(y)\right)\\ &= a^2\varphi_1(x)\varphi_1(y) + b^2\varphi_2(x)\varphi_2(y)\\ &= 5\varphi_1(x)\varphi_1(y) + 4\varphi_2(x)\varphi_2(y)\end{aligned}$$

לכן נבחר ש- $a = \sqrt{5}, b = \sqrt{4}$ ונקבל ש-

$$\begin{aligned}\varphi_3(x)\varphi_3(y) &= \left(\sqrt{5}\varphi_1(x), \sqrt{4}\varphi_2(x)\right)\left(\sqrt{5}\varphi_1(y), \sqrt{4}\varphi_2(y)\right)\\ &= 5\varphi_1(x)\varphi_1(y) + 4\varphi_2(x)\varphi_2(y) = 5K_1 + 4K_2 = K\end{aligned}$$

כלומר, מצאנו שהמיפוי $\varphi_3(x) = \left(\sqrt{5}\varphi_1(x), \sqrt{4}\varphi_2(y)\right)$ היא פונקציית המיפוי של K.

■

1. b) מאחר ונתון שקיים מפריד לינארי ב- R^m עם פונקציית המיפוי φ_1 מ- R^n ל- R^m אזי קיים וקטור משקלות $w=(w_1,\dots,w_m)$ שהוא וקטור המשקולות של המפריד הלינארי וש- $sign\left(w\varphi_1(x)\right)$ הוא מפריד לינארי ב- R^m. מתקיים כי $\varphi_1(x)=\varphi_1(x_1,\dots,x_n)=(x'_1,\dots,x'_m)$. נסמן ב- k את המימד שאליו ממפה פונקציית המיפוי φ_2.נראה כי קרנל K מסעיף $1.a$ ממפה למימד $m+k$. נסמן את וקטור המשקולות $w'=(w'_1,\dots,w'_m,\dots,w'_{m+k})$ ומסעיף $1.a$ נראה כי $\varphi_3(x)=\left(\sqrt{5}\varphi_1(x),\sqrt{4}\varphi_2(x)\right)$. נראה כי עבור $w'_i=0$, $m+1\le i\le m+k$ נקבל כי

$$sign\left(w'\varphi_3(x)\right)=sign\left((w'_1,\dots' w'_{m+k})\left(\sqrt{5}\varphi_1(x),\sqrt{4}\varphi_2(x)\right)\right)=$$

$$sign\left(\sqrt{5}w'_1x'_1+\dots+\sqrt{5}w'_mx'_m+\underbrace{0+\dots+0}_{k}\right)=$$

$sign\left(\sqrt{5}w'_1x'_1+\dots+\sqrt{5}w'_mx'_m\right)$ ונראה כי עבור $w'_i=\frac{1}{\sqrt{5}}w_i$, $1\le i\le m$ נקבל ש-

$$sign\left(\sqrt{5}w'_1x'_1+\dots+\sqrt{5}w'_mx'_m\right)=$$

$$sign\left(\sqrt{5}*\frac{1}{\sqrt{5}}w_1x'_1+\dots+\sqrt{5}*\frac{1}{\sqrt{5}}w_mx'_m\right)=sign(w_1x'_1+\dots+w_mx'_m)$$
$$=sign(w\varphi_1(x))$$

ואנו יודעים כי $sign(w\varphi_1(x))$ הוא מפריד לינארי במימד m ולכן נקבל כי $sign\left(w'\varphi_3(x)\right)$ הוא מפריד לינארי במימד $m+k$ לפי קרנל K כך ש- $w'=(\frac{1}{\sqrt{5}}w_1,\dots,\frac{1}{\sqrt{5}}w_m,\underbrace{0,\dots,0}_{k})$. ■

1. c) נגדיר $\varphi:R\to R^N\ s.t\ \varphi(x)=(\underbrace{3,\dots,3}_{x\ times},\underbrace{0,\dots,0}_{N-x\ times})$. יהי $x,y\in S$ ונראה כי

$$\varphi(x)*\varphi(y)=\left(\underbrace{3,\dots,3}_{x},\underbrace{0,\dots,0}_{N-x}\right)\left(\underbrace{3,\dots,3}_{y},\underbrace{0,\dots,0}_{N-y}\right)=$$
$$=\underbrace{3*3+\dots+3*3}_{\min(x,y)}+0\dots+0=9*\min(x,y)$$

אזי מצאנו כי K הוא קרנל כך ש- $K=\varphi(x)\varphi(y)$ ופונקציית המיפוי המתאימה φ היא $\varphi(x)=\left(\underbrace{3,\dots,3}_{x},\underbrace{0,\dots,0}_{N-x}\right)$. ■

2. Lagrange multipliers (20 pts)

Suppose you are running a factory, producing some sort of widget that requires steel as a raw material. Your costs are predominantly human labor, which is \$20 per hour for your workers, and the steel itself, which runs for \$170 per ton.

Suppose your revenue R is modeled by the following equation:

$$R(h,s) = 200 \cdot h^{\frac{2}{3}} \cdot s^{\frac{1}{3}}$$

Where:

- h represents hours of labor
- s represents tons of steel

If your budget is \$20,000, what is the maximum possible revenue?

תשובה:

2. נגדיר $G(h,s) = 20h + 170s = 20{,}000$ שזה סכום ההוצאות, $R(h,s) = 200h^{\frac{2}{3}} * s^{\frac{1}{3}}$ וזה פונקציית ההכנסות. יהי $L(h,s) = 200h^{\frac{2}{3}} * s^{\frac{1}{3}} + \lambda(20h + 170s - 20{,}000)$. נראה כי

$$\frac{\partial L}{\partial h} = \frac{2}{3} * 200 * h^{-\frac{1}{3}} * s^{\frac{1}{3}} + 20\lambda = 0 \rightarrow \frac{400}{3} * s^{\frac{1}{3}} * h^{-\frac{1}{3}} = -20\lambda \ (*)$$

$$\frac{\partial L}{\partial s} = \frac{1}{3} * 200 * h^{\frac{1}{3}} * s^{-\frac{2}{3}} + 170\lambda = 0 \rightarrow \frac{200}{3} * h^{\frac{2}{3}} * s^{-\frac{2}{3}} = -170\lambda \ (**)$$

$$\frac{*}{**} \rightarrow 2 * \frac{s}{h} = \frac{2}{17} \rightarrow 17s = h \ (*)$$

$$\frac{\partial L}{\partial \lambda} 20h + 170s - 20{,}000 = 0 \rightarrow_{*} 20 * 17s + 170s - 20{,}000 = 0 \rightarrow$$

$$510s = 20{,}000 \rightarrow \boldsymbol{s} = \boldsymbol{\frac{20{,}000}{510}}$$

$$h = 17s = 17 * \frac{20{,}000}{510} = \boldsymbol{\frac{20{,}000}{30}}$$

$$170\lambda = \frac{200}{3} * h^{\frac{2}{3}} * s^{-\frac{2}{3}} \rightarrow \lambda = \frac{\frac{200}{3} * \left(\frac{20{,}000}{30}\right)^{\frac{2}{3}} * \left(\frac{20{,}000}{510}\right)^{-\frac{2}{3}}}{170} = \mathbf{2.59}$$

אזי מצאנו כי פונקציית ההכנסות המקסימלית היא $R(h,s) = 200h^{\frac{2}{3}} * s^{\frac{1}{3}}$ כך ש- $R\left(\frac{20{,}000}{510}, \frac{20{,}000}{30}\right) = 200 * \left(\frac{20{,}000}{30}\right)^{\frac{2}{3}} * \left(\frac{20{,}000}{510}\right)^{\frac{1}{3}} = \mathbf{51854.51\$}$ וכאשר נחסר את ההוצאות אז נקבל כי סכום הרווח המקסימלי הוא $51854.51 - 20{,}000 = \mathbf{31{,}854.51\$}$. ■

3. <u>PAC Learning and VC dimension (30 pts)</u>

Let $X = \mathbb{R}^2$. Let

$$C = H = \left\{ h(r_1, r_2) = \left\{ (x_1, x_2) \middle| \begin{matrix} x_1^2 + x_2^2 \geq r_1 \\ x_1^2 + x_2^2 \leq r_2 \end{matrix} \right\} \right\}, \text{ for } 0 \leq r_1 \leq r_2,$$

the set of all origin-centered rings.

a. (8 pts) What is the $VC(H)$? Prove your answer.
b. (14 pts) Describe a polynomial sample complexity algorithm L that learns C using H. State the time complexity and the sample complexity of your suggested algorithm. Prove all your steps.

In class we saw a bound on the sample complexity when H is finite.

$$m \geq \frac{1}{\varepsilon}\left(\ln|H| + \ln\frac{1}{\delta}\right)$$

When $|H|$ is infinite, we have a different bound:

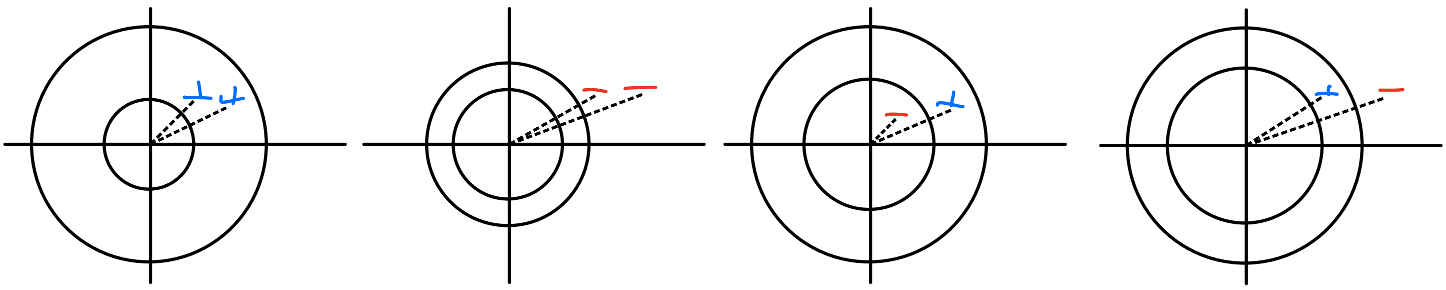
$$m \geq \frac{1}{\varepsilon}\left(4\log_2\frac{2}{\delta} + 8VC(H)\log_2\frac{13}{\varepsilon}\right)$$

c. (8 pts) You want to get with 95% confidence a hypothesis with at most 5% error. Calculate the sample complexity with the bound that you found in b and the above bound for infinite $|H|$. In which one did you get a smaller m? Explain.

<u>תשובה:</u>

a) נוכיח כי $VC(H) = 2$. נשים לב כי H מרחב ההיפותזות כך שכל $h \in H$ הוא "טבעת" כי שני מעגלים כאשר האחד מוכל בשני יוצר מעין טבעת. נסמן את הנקודות שנמצאות בתוך הטבעת ב- + ובכחול ואת מה שמחוץ לטבעת ב- - (מינוס) ובאדום.

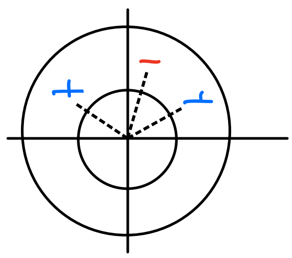
נראה כי $VC(H) \geq 2$– נראה כי עבור קבוצה של 2 דוגמיות כך שלכל דיכוטומיה עבורם קיים היפותזה $h \in H$ אשר מנתצת אותם. נבחר 2 דגימות אשר הן בקו לינארי (כלומר, המרחק ביניהן הוא קו ישר) והקו מקביל לציר x ולכן יש $2^2 = 4$ דיכוטומיות ונראה שלכל דיכוטומיה קיים $h \in H$ שקונסיסטנטית.



ולכן $VC(H) \geq 2$.

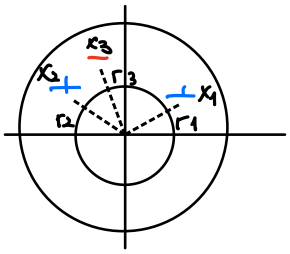
כעת, נראה כי $VC(H) < 3$- לפי מה שראינו צריך להראות קבוצה של 3 דגימות כך שקיימת דיכוטומיה אחת שלא קיימת לה $h \in H$ אשר קונסיסטנטית איתה. נראה כי ישנם 2 מקרים בלבד לחלוקת 3 נקודות במרחב-

- 3 הנקודות באותו הרדיוס מראשית הצירים ונבחר את הדיכוטומיה-



נשים לב כי מאחר והדגימות הן באותו רדיוס מראשית הצירים אז כל טבעת שתכיל את אחת מהדגימות תכיל גם את האחרות ולכן בדיכוטומיה זאת לא קיימת היפותזה שתנתץ אותם.

- קיימת דגימה אחת לפחות שהרדיוס שלה שונה מהרדיוס של הדגימות האחרות ונראה את הדיכוטומיה-



נראה כי r_2, r_3 הם רדיוסים שווים באורכם אבל $r_2, r_3 > r_1$ ונראה ש- x_1, x_2 התגית שלהם היא + אז הטבעת מכילה את שניהם וגם בהכרח צריכה להכיל את x_3 כי $r_2 = r_3$.

לכן הראנו כי $VC(H) < 3$ ולכן הוכחנו ש- $VC(H) = 2$. ■

b) האלגוריתם הוא:

- נאתחל את $r_1, r_2 = 0$.
- עבור כל דגימה שבקבוצה- $\{(x_1, x_2) \in D | C(x_1, x_2) = true\}$ נחשב את $x_1^2 + x_2^2$.
- אם קיבלנו ש- $x_1^2 + x_2^2 > r_2$ אז נעדכן כי $x_1^2 + x_2^2 = r_2$.
- אם קיבלנו ש0- $x_1^2 + x_2^2 < r_1$ אז נעדכן כי $x_1^2 + x_2^2 = r_1$.

בסוף מחזירים את הטבעת שהרדיוס החיצוני הוא r_2 והרדיוס הפנימי הוא r_1.

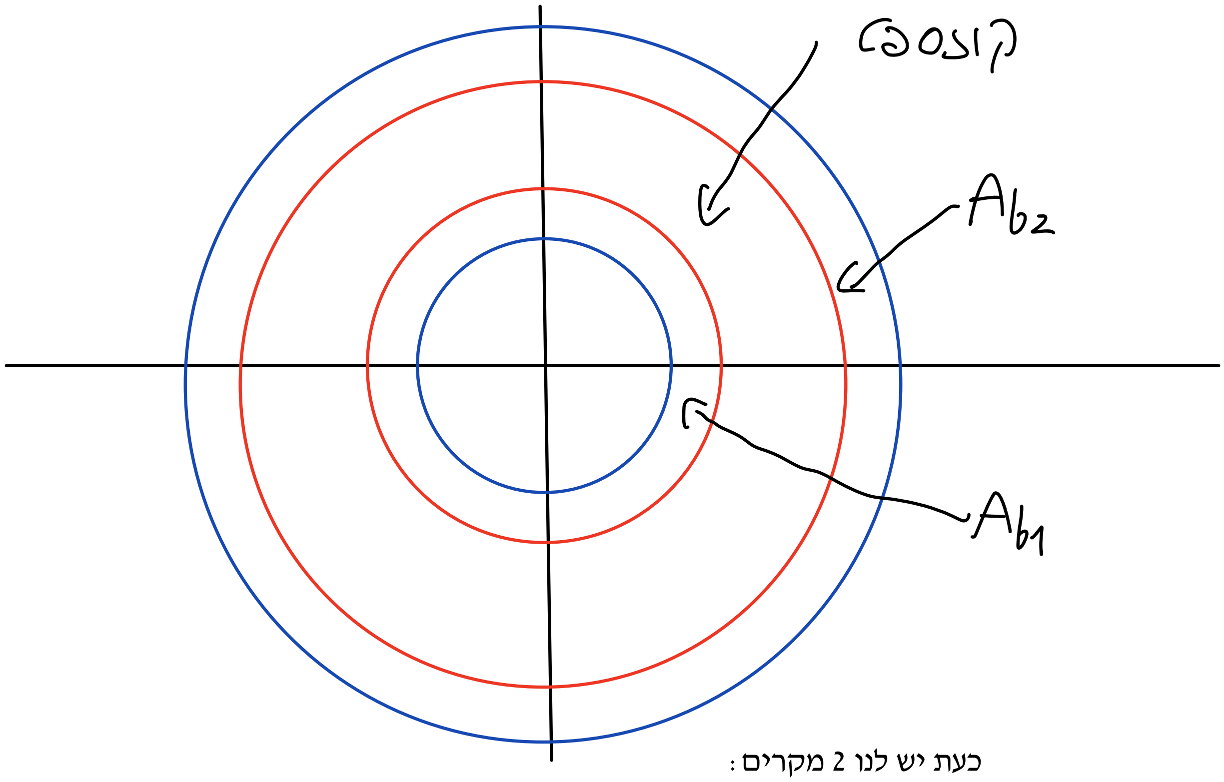
<u>נוכיח כי האלגוריתם consistent עם המידע</u>: מאחר ואתחלנו את $r_1, r_2 = 0$ אז אנו עוברים על כל הדגימות שהתגית שלהן הן + ואז אנו מגדילים/מקטינים את r_2, r_1 בהתאמה לפי הדגימות. זה יוצר לנו 2 מעגלים אשר מרכז המעגל שלהם זהה ושניהם יוצרים טבעת. כך בעצם כל דגימה חיובית תופיע בתוך הטבעת וכל דגימה שלילית לא תופיע בתוך הטבעת.

<u>סיבוכיות דגימה</u>: יהיו $\varepsilon > 0, \delta > 0$. עבור $c \in C$ נגדיר את r_1, r_2 להיות הרדיוס החיצוני והרדיוס הפנימי של הטבעת בהתאמה. נגדיר את c^ε להיות הטבעת הנוצרת מכיווץ $\frac{\varepsilon}{2}$ מהשטח של הטבעת החיצונית המוגדרת על ידי r_2 וכיווץ $\frac{\varepsilon}{2}$ מהשטח של הטבעת הפנימית המוגדרת על ידי r_1. נגדיר עבור $r_1 \le b \le r_2$ ש-

$$A_{b_1} = \{(x_1, x_2) | b \le d\big((x_1, x_2), (0,0)\big) \le r_2$$

$$A_{b_2} = \{(x_1, x_2) | r_1 \le d\big((x_1, x_2), (0,0)\big) \le b$$

ונגדיר $r_2^\varepsilon = \inf\{b | \pi\big(A_{b_2}\big) \le \frac{\varepsilon}{2}\}, r_1^\varepsilon = sup\{b | \pi(A_{b_1}) \le \frac{\varepsilon}{2}\}$.



כעת יש לנו 2 מקרים :

- המקרה הטוב- קיימת לפחות דגימה אחת שנמצאת ב- A_{b_1} או ב- A_{b_2}. נראה כי ההסתברות שמקרה כזה יקרה היא
$$1-\left(1-\frac{\varepsilon}{2}\right)^m+\left(1-\frac{\varepsilon}{2}\right)^m=1-2\left(1-\frac{\varepsilon}{2}\right)^m<\varepsilon$$
ולכן מקרה זה לא מעניין אותנו.
- המקרה הרע- כל הדגימות לא נופלות ב- A_{b_1} וב- A_{b_2}. נראה כי ההסתברות שמקרה כזה יקרה היא-
$$\left(1-\frac{\varepsilon}{2}\right)^m+\left(1-\frac{\varepsilon}{2}\right)^m=2\left(1-\frac{\varepsilon}{2}\right)^m\leq 2\varepsilon^{-\frac{\varepsilon m}{2}}$$

מאחר והדגימות הן בלתי תלויות אז נראה כי
$$2\varepsilon^{-\frac{\varepsilon m}{2}}<\delta\to\varepsilon^{-\frac{\varepsilon m}{2}}<\frac{\delta}{2}\to\ -\frac{\varepsilon m}{2}<\ln\left(\frac{\delta}{2}\right)\to m\geq\frac{-\ln\left(\frac{\delta}{2}\right)*2}{\varepsilon}$$
$$=\frac{\ln\left(\frac{\delta}{2}\right)*2}{\varepsilon}$$

<u>סיבוכיות זמן :</u> נראה כי אתחול r_1,r_2 לוקח $O(1)$. ביצוע חישוב $r_1^2+r_2^2$ בעבור כל דגימה הוא $O(1)*m=O(m)$. החזרת r_1,r_2 מתאימים לוקח $O(1)$. לכן סה"כ נקבל שזמן הריצה הוא $O(1)+O(m)+O(1)\approx$ $O(m)$. ■

c) נחשב לפי הגבול שמצאנו בסעיף b – נציב $\varepsilon=0.05$, $\delta=0.95$ ונראה כי

$m\geq\frac{\ln\left(\frac{2}{0.05}\right)*2}{0.05}=147.5\to 148_{\text{דגימות}}$ וכעת נחשב לפי הנוסחה ש- $|H|$ אינסופי-

$$m\geq\frac{1}{\varepsilon}(4\log_2\left(\frac{2}{\delta}\right)+8VC(H)\log_2\left(\frac{13}{\varepsilon}\right))$$
$$=\frac{1}{0.05}(4\log_2\left(\frac{2}{0.05}\right)+8*2*\log_2\left(\frac{13}{0.05}\right))=2992.912$$
$$\to\ 2993_{\text{דגימות}}$$

קיבלנו חסם הדוק יותר בנוסחה מסעיף b מאחר והנוסחה של $|H|$ האינסופית לא מניחה שום הנחה על המידע מלבד ה- $VC(H)$ לעומת החסם שמצאנו ב- b שמתחשב במידע על H שעוזר לנו לבצע חישוב יותר מדויק. ■

4. <u>VC dimension (20 pts)</u>

Let $X = \mathbb{R}$ and $n \in \mathbb{N}$.

Define "x-node decision tree" for any $x = 2^n - 1$ to be a full binary decision tree with x nodes (including the leaves).

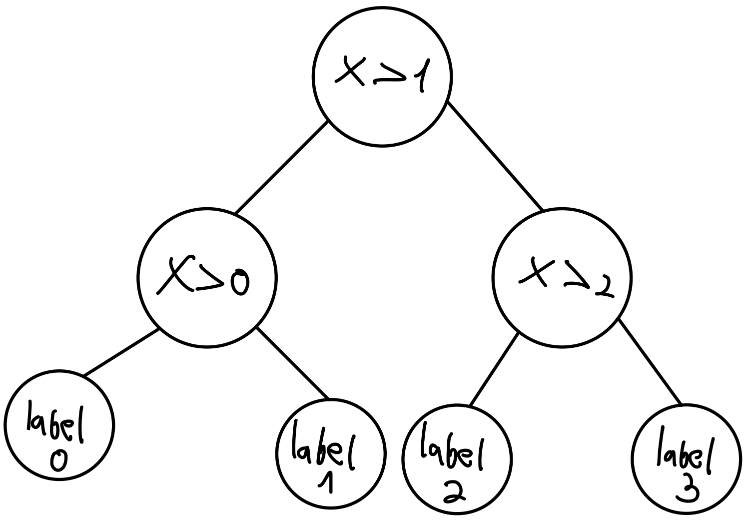
Let H_m be the hypothesis space of all "x-node decision tree" with $n \leq m$.

a. (5 pts) What is the $VC(H_3)$? Prove your answer.

b. (15 pts) What is the $VC(H_m)$? Prove your answer.

<u>תשובה:</u>

a) נראה כי עבור H_3 אז זה מרחב ההיפותזות כך ש- $n > 3$ ולכן מספר ה- $nodes$ האפשריים עבור עצים במרחב זה הוא $2^1 - 1 = 1, 2^2 - 1 = 3, 2^3 - 1 = 7$. נוכיח כי $VC(H_3) = 4$.
נראה כי $VC(H_3) \geq 4$- נבחר את הקבוצה $\{0,1,2,3\}$ ונראה כי



נשים לב כי עבור כל דיכוטומיה על הקבוצה נוכל לבחור $h \in H_3$ על אותו העץ חוץ מהתגיות של העלים. ניתן לראות כי בכל דגימה אנו הולכים לעלה שונה שזה מראה שניתן לשנות את התגית של

העלים בהתאם לתגיות של הדגימות ולכן ההיפותזה תמיד תהיה קונסיסטנטית עם הקבוצה. לכן H_3 מנתצת קבוצה של 4 איברים.

כעת נוכיח כי $VC(H_3) < 5$- תהי הקבוצה $x = \{x_1, x_2, x_3, x_4, x_5\}$ ב- R כך ש-

$x_1 \le x_2 \le x_3 \le x_4 \le x_5$. נבחר את הדיכוטומיה-

$c(x_1) = 0, c(x_2) = 1, c(x_3) = 0, c(x_4) = 1, c(x_5) = 0$. נראה כי לא ניתן לנתץ דיכוטומיה זאת ולכן לא קיימת $h \in H_3$ כך ש- $h(x_i) = c(x_i), s.t\ 1 \le i \le 5$. מהיות ועבור

$h \in H_3$ אז העץ עם מס' הצמתים הגדול ביותר הוא העץ שמס' הצמתים בו הוא 7 ומהיות והעץ הוא עץ חיפוש בינארי מלא אז מס העלים בעץ זה הוא 4. מהיות וב- x יש 5 דגימות אז לפי עקרון שובך היוני קיימות 2 דגימות שהולכות לאותו עלה. לכן, נניח בלי הגבלת הכלליות כי הנקודות הן $x_i \le x_j$. נחלק למקרים:

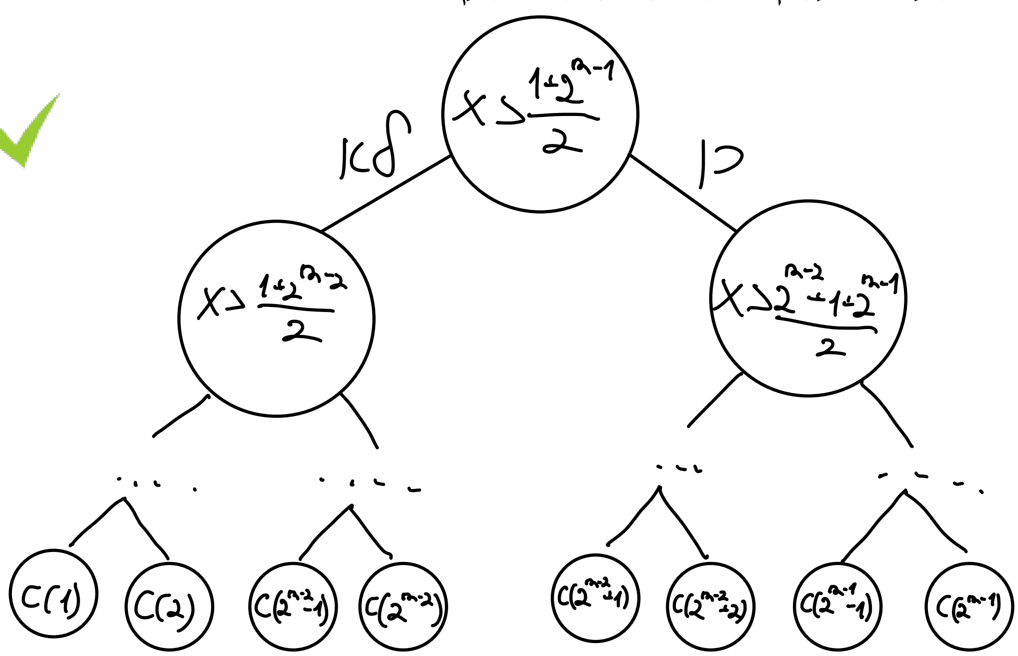
1. אם x_i, x_j הן דגימות עוקבות אז $c(x_i) \ne c(x_j)$ בהכרח לפי בחירת הדיכוטומיה שלנו אבל הם סווגו לאות עלה, כלומר $h(x_i) = h(x_j)$.
2. אחרת, אם x_i, x_j הן לא דגימות עוקבות אז קיים x_{i+1} כך ש- $x_i \le x_{i+1} \le x_j$. מהיות ו- x_i, x_j סווגו לאותו עלה ומהיות שהדגימות הן מ- R אז לפי עץ ההחלטה בהכרח גם x_{i+1} מסווג לאותו עלה, כלומר מתקיים $h(x_i) = h(x_{i+1})$ וזו טעות מאחר ולפי הקונספט שהגדרנו אז $c(x_i) \ne c(x_{i+1})$.

לכן, קיבלנו כי $VC(H_3) = 4$. ■

b) מהיות ובעץ בינארי מלא עם $2^m - 1$ צמתים יש 2^{m-1} עלים אז נוכיח כי $VC(H_m) = 2^{m-1}$.
נראה כי $VC(H_m) \ge 2^{m-1}$- מהיות והעץ עם מס' הצמתים הגדול ביותר הוא העץ עם מס' העלים הגדול ביותר אז נסתכל על העץ עם $2^m - 1$ צמתים. נבחר קבוצה עם 2^m דגימות

$S = \{1, \dots, 2^m\}$. נראה כי לכל עץ החלטה עם קונספט אחר התבנית הבאה זהה-

כל צומת מייצגת טווח מסוים כך שהטווח הוא בעצם הוא הדגימה המינימלית מבין הדגימות שנותרו והדגימה המקסימלית מבין הדגימות שנותרו. המעבר בין הצמתים הוא שאם $x \in S$ מקיי ש- $x > \frac{\min(instance\ left) + \max(instance\ left)}{2}$ אז הולכים ימינה, אחרת הולכים שמאלה. בעצם כך עבור 2^{m-1} העלים יגיע רק דגימה אחת. נראה את העץ-



נראה כי עבור כל קונספט על הקבוצה ניתן לבחור $h \in H_m$ והעץ יישאר זהה חוץ מהתגיות של העלים. כמו שניתן לראות כל דוגמית מסווגת לעלה שונה, כלומר ניתן לשנות את התגיות של העלים בהתאם לתגיות של הדוגמיות. לכן, H_m מנתצת את הקבוצה עם 2^{m-1} איברים ולכן $VC(H_m) \geq 2^{m-1}$.

נראה כעת כי $VC(H_m) < 2^{m-1} + 1$- יהי $X = \{x_1, \dots, x_{2^{m-1}+1}\}$ כך ש- $|X| = 2^{m-1} + 1$ ומתקיים כי $x_1 \leq x_2 \leq \cdots \leq x_{2^{m-1}+1}$. נבחר את הקונספט הבא-

$$c(x_1) = 0, c(x_2) = 1, \dots, c(x_{2^{m-1}+1}) = 0$$

כלומר עבור i זוגי אז הקונספט של הדגימה הוא 1 ואחרת הוא 0. כעת מהיות ו- h הוא עץ בינארי מלא אז יש לכל היותר 2^{m-1} עלים. לכן, לפי עקרון שובך היונים בהכרח יש 2 דגימות המסווגות לאותו עלה. נסמן את הדגימות האלה כך $x_i \leq x_j$. נחלק למקרים:

- אם x_i, x_j הן דגימות עוקבות אז מהגדרת הקונספט נקבל כי $c(x_i) \neq c(x_j)$ אבל מתקיים ש- $h(x_i) = h(x_j)$ ולכן h לא קונסיסטנטית.
- אם x_i, x_j הן לא דגימות עוקבות אז קיים $x_i \leq x_{i+1} \leq x_j$ כאשר שלושתם מסווגים לאותו עלה ב- h בהכרח. לכן מתקיים כי $h(x_i) = h(x_{i+1})$ אבל לפי הקונספט שהגדרנו $c(x_i) \neq c(x_{i+1})$. לכן h לא קונסיסטנטית.

לכן, לא קיימת היפותזה שמסווגת בצורת קונסיסטנטית את הדיכוטומיה הזאת. לכן, לא ניתן לנתץ $2^{m-1} + 1$ נקודות ב- H_m.

הראינו כי $2^{m-1} \leq VC(H_m) < 2^{m-1} + 1$ ולכן נקבל כי $VC(H_m) = 2^{m-1}$. ■