HW5 – Theory + SVM

1. PAC Learning and VC dimension (30 pts)

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

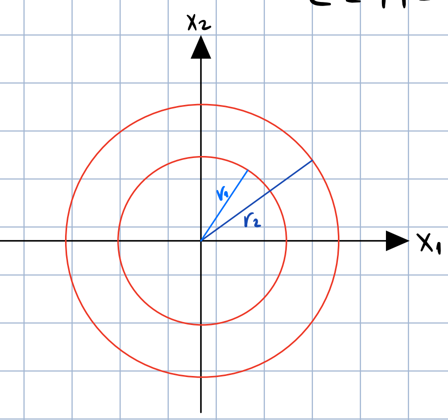
, for $0 \leq r_1 \leq r_2$,

the set of all origin-centered rings.

a. (8 pts) What is the $VC(H)$? Prove your answer.

$$C = H = \Big\{ \underbrace{h(r_1,r_2)}_{\text{כל הטבעות שהרדיוסים האלה יוצרים}} = \underbrace{\left\{ (x_1,x_2) \,\middle|\, \begin{matrix} x_1^2 + x_2^2 \geq r_1 \\ x_1^2 + x_2^2 \leq r_2 \end{matrix} \right\}}_{\text{6 הנקודות בתוך הטבעת}} \Big\}, \quad \text{for } \underbrace{0 \leq r_1 \leq r_2}$$

r_2 הוא הגדול יותר, לכן r_2 הוא הרדיוס החיצוני



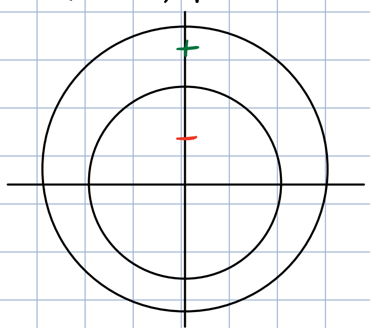
* $H =$ כל 6 הטבעות שמרכזן בראשית הצירים
* $H=C$ – מרחב ההיפותזה שווה למרחב הקונספט (זו הנחה עיקרית)

a) ראשית, נראה ש $VC(H)\geq 2$:

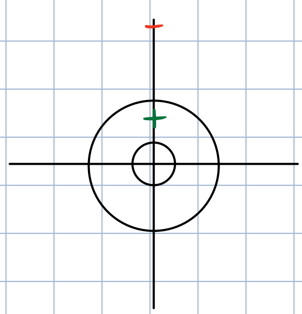
נבחר שתי נקודות p_1,p_2 כך ש $p_1=(0,1)$ ו- $p_2=(0,3)$

כעת נבחן את ארבעת הקונספטים האפשריים:

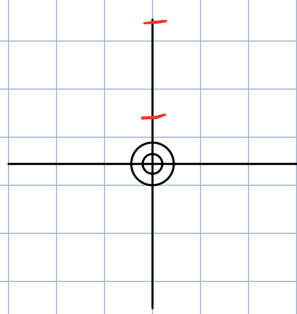
$p_1 = -\ ,\ p_2 = +$



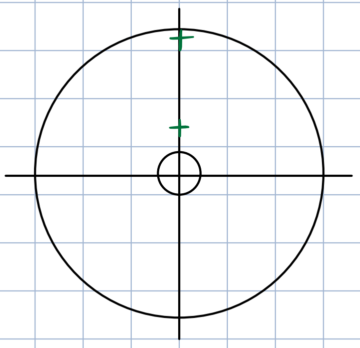
$p_1 = +\ ,\ p_2 = -$



$p_1 = -\ ,\ p_2 = -$



$p_1 = +\ ,\ p_2 = +$



עכשיו נראה ש $VC(H) < 3$:

מרכז הטבעות הוא בראשית הצירים, לכן נפריד למקרים לפי המרחק של הנקודות ממנו -

1) אם שלושת ה-x במרחק שונה מראשית הצירים, במקרה בו הנקודה הרחוקה ביותר והנקודה הקרובה ביותר מסווגות כ + (בתוך הטבעת) והנקודה האמצעית כ - (בחוץ), אין דרך לבצע Shattering.

2) אם 2 נקודות באותו מרחק מראשית הצירים, אז לא יהיה ניתן להפריד ביניהן אף פעם (בדוגמא שבה אחת צריכה להיות בתוך הטבעת והאחרת בחוץ נכשל בניסוי).

3) אם כולן באותו מרחק אז ברור כי לא ניתן להפריד בצורה כזאת (הסבר דומה למקרה 2).

לכן $VC(H)=2$.

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:

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

(b נגדיר את האלגוריתם L באופן הבא-

for every p in m (given data points):

Calculate the distance from p to origin

d_1 = minimum distance

d_2 = maximum distance

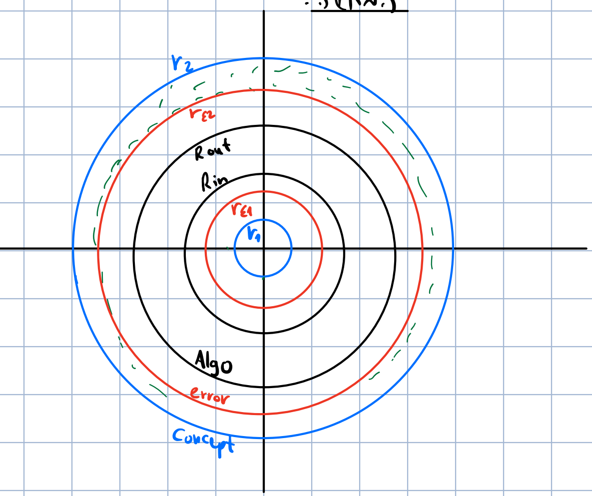
$R_{in} = d_1$

$R_{out} = d_2$

$P =$ הנקודות שנתונות לנו (Training Data)

$m =$ כל הנקודות ב Data

המחשה:



כעת נחשב את ה Sample complexity:

נגדיר $r_{\varepsilon 1}$ ואת $r_{\varepsilon 2}$ כך ש $r_1 \leq r_{\varepsilon 1} \leq r_{\varepsilon 2} \leq r_2$ כך ש,

$$A_1 = \{(x_1, x_2) \mid r_1 \leq \sqrt{x_1^2 + x_2^2} \leq r_{\varepsilon 1}\}$$

$$A_2 = \{(x_1, x_2) \mid r_{\varepsilon 2} \leq \sqrt{x_1^2 + x_2^2} \leq r_2\}$$

כך ש, $\pi(A_1) = \frac{\varepsilon}{2}$, $\pi(A_2) = \frac{\varepsilon}{2}$

נחלק למקרים:

1) $r_{\varepsilon 2} \leq r_{out} \leq r_2$, $r_1 \leq r_{in} \leq r_{\varepsilon 1}$:

במקרה זה $\pi(x) \leq \varepsilon$ עבור $x \in R^d$ ולכן מקרה זה

גם מטופל.

2) $r_{\varepsilon2} \le r_{out} \le r_2$ וגם $r_1 \le r_{\varepsilon1} < r_{in}$ או $r_{out} < r_{\varepsilon2} \le r_2$, $r_1 \le r_{in} \le r_{\varepsilon1}$:

במקרה זה לבחור נקודה אחת, נמצא במרחב שבין המלבן של $r_{\varepsilon1}$ או $r_{\varepsilon2}$ לבין המלבנים של r_1 ו- r_2 בהתאמה.

הסיכוי לטעות סיווג במקרה זה הוא לכל היותר $\frac{\varepsilon}{2}\cdot\left(1-\frac{\varepsilon}{2}\right)^{m-1} \le \frac{\varepsilon}{2} < \varepsilon$ ולכן גם מקרה זה לא מעניין.

3. $r_1 \le r_{\varepsilon1} < r_{in}$ וגם $r_{out} < r_{\varepsilon2} \le r_2$:

במקרה זה הסיכוי לטעות הוא גדול מ- ε, בפרט, הסיכוי הוא $(1-\varepsilon)^m \le e^{-m\varepsilon}$

אם Training Data בגודל $\frac{\ln(\frac{1}{\delta})}{\varepsilon} \le m$ נקבל שהסיכוי לקבלת מקרה כזה הוא

$$(1-\varepsilon)^m \le e^{-m\cdot\varepsilon} \le e^{-\frac{\ln(\frac{1}{\delta})}{\varepsilon}\cdot\varepsilon} = e^{-\ln(\frac{1}{\delta})} = \delta$$

כעת נחשב את הסיבוכיות זמן

מכיוון שבכל תהליך אימון למציאת היפותזה אנו בודקים כל instance פעם אחת

הסיבוכיות היא $O(n)$

c) עבור החסם שמצאנו בסעיף b:

$\delta = 0.95$, $\varepsilon = 0.05$,
$$m \ge \frac{\ln(\frac{1}{\delta})}{\varepsilon} = \frac{\ln(\frac{1}{0.95})}{0.05} \cong 1.03$$

לכן $m \ge 2$.

עבור החסם שמצאנו בכיתה ל $|H|$ אינסופי:

$$m \ge \frac{1}{\varepsilon}\left(4\cdot\log_2\frac{2}{\delta} + 8\cdot VC(H)\cdot\log_2\left(\frac{13}{\varepsilon}\right)\right) = \frac{1}{0.05}\left(4\cdot\log\frac{2}{0.95} + 8\cdot 2\cdot\log_2\frac{13}{0.05}\right) \cong 2653.078$$

ולכן $m \ge 2654$

קיבלנו שהחסם שאנחנו בנינו הדוק יותר וזה מכיוון שבנינו מקרה פרטי עבור c הספציפי שלנו

ואילו החסם שמצאנו בכיתה הוא חסם כללי עבור כל c.

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

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

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

$x = 2^n - 1$

Let H_m be the hypothesis space of all “x-node decision tree” with $n \le m$.

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

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

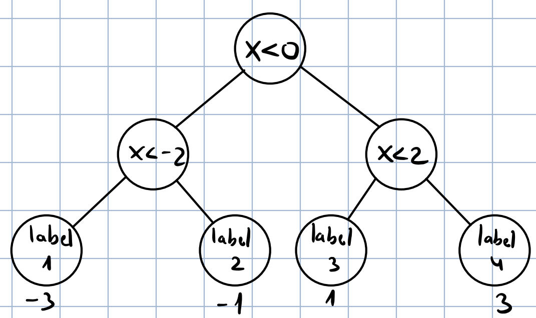
②

a) מהגדרה, H_3 הוא מרחב ההיפותזות כך ש $n \le 3$. לכן יש 3 אפשרויות:

$n=0$, 2^0-1 , 0 nodes
$n=1$, 2^1-1 , 1 nodes
$n=2$, 2^2-1 , 3 nodes
$n=3$, 2^3-1 , 7 nodes

ראשית, נוכיח כי $VC(H_3) \ge 4$:

מכיוון $x=4$ נבחר את הקבוצה $\{-3, -1, 1, 3\}$.
עבור העץ הבינארי המלא הבא -



* ברור כי עצים שבהם $n \in \{0,1,2\}$ מוכלים ב H_3
ולכן נראה את המקרה שבו $n=3$.

בעץ כזה העלים ניתנים להפרדה כך שכל x נמצא בעלה נפרד.
ולכן כל קומבינציה שנקבל ניתן לשנות את צמתי העץ ברמות הפנימיות ולכן ההיפותזה תהיה קונסיסטנטית עם הקבוצה.

ולכן ב- H_3 ניתן לעשות shattering ל- 4 איברים.

כעת, נראה כי $VC(H_3) < 5$:

בהינתן 5 נקודות, מעיקרון שובך היונים בהכרח לפחות בעלה אחד יהיו לפחות 2 instances, בכל עץ ב H_3.
לכן קיימת לפחות קומבינציה אחת שבה העץ לא יוכל להפריד, מכיוון שאם בעלה זה (שבו 2 instances)
אחד מהם יסווג ל + והשני ל - לא יהיה ניתן להפריד ביניהם בתהליך הניתוח.

לכן, $VC(H_3)=4$

b) ראשית, נראה כי $VC(H_m) \geq 2^{m-1}$:

בעץ בינארי מלא עם $2^m - 1$ צמתים יש 2^{m-1} עלים.

לכן נבחר 2^{m-1} נקודות - $p_1, p_2, \dots p_{2^{m-1}}$ כך ש $p_1 < p_2 < \dots < p_{2^{m-1}}$, ונבנה עץ שבו תהיה נקודה אחת בכל עלה (בדומה לסעיף a).

כך בכל דיכוטומיה כל עלה יכול לבחור את הערך הרצוי, בדומה להסבר בסעיף a.

כעת, $VC(H_m) < 2^{m-1}+1$:

יהיו $2^{m-1}+1$ נקודות: $p_1, p_2, \dots, p_{2^{m-1}+1}$. בדומה לכתוב לעיל, ראינו שבעץ כזה ישנם 2^{m-1} עלים.

לכן מעיקרון שובך היונים, בכל עץ ששייך ל H_m, יהיו לפחות 2 נקודות בלפחות עלה 1.

נבחר דיכוטומיה שבה שתי נקודות אלה מסווגות כאחת + והשניה -, ונראה שלא ניתן לעץ מספר כזה של ערכים.

לכן קיבלנו ש $VC(H_m) = 2^{m-1}$

3. Kernels and mapping functions (25 pts)

a. (20 pts) Let $K(x,y) = (x \cdot y + 1)^3$ be a function over $\mathbb{R}^2 \times \mathbb{R}^2$ (i.e., $x, y \in \mathbb{R}^2$).

Find ψ for which K is a kernel. (It may help to first expand the above term on the right-hand side).

3

a) נפתח את הצד הימני:

$$(\overset{a}{x \cdot y} + \overset{b}{1})^3 = a^3 + 3a^2b + 3ab^2 + b^3$$

$$= xy^3 + 3(xy)^2 \cdot 1 + 3 \cdot xy \cdot 1 + 1$$

$$= (xy)^3 + 3 \cdot (xy)^2 + 3 \cdot xy + 1$$

$$= (x_1y_1 + x_2y_2)^3 + 3 \cdot (x_1y_1 + x_2y_2)^2 + 3(x_1y_1 + x_2y_2) + 1$$

$$= (x_1y_1)^3 + 3 \cdot (x_1y_1)^2(x_2y_2) + 3(x_1y_1)(x_2y_2)^2 + (x_2y_2)^3 + 3\left((x_1y_1)^2 + 2(x_1y_1)(x_2y_2) + (x_2y_2)^2\right) + 3(x_1y_1 + x_2y_2) + 1$$

$$= (x_1y_1)^3 + 3 \cdot (x_1y_1)^2(x_2y_2) + 3(x_1y_1)(x_2y_2)^2 + (x_2y_2)^3 + 3(x_1y_1)^2 + 6(x_1y_1)(x_2y_2) + 3(x_2y_2)^2 + 3(x_1y_1) + 3(x_2y_2) + 1$$

$$= x_1^3y_1^3 + 3 \cdot x_1^2y_1^2 \cdot x_2 \cdot y_2 + 3 \cdot x_1 \cdot y_1 \cdot x_2^2 \cdot y_2^2 + x_2^3 \cdot y_2^3 + 3 \cdot x_1^2 \cdot y_1^2 + 6 \cdot x_1 \cdot y_1 \cdot x_2 \cdot y_2 + 3 \cdot x_2^2 \cdot y_2^2 + 3 \cdot x_1 \cdot y_1 + 3 \cdot x_2 \cdot y_2 + 1$$

$$= (x_1^3 + \sqrt{3} \cdot x_1^2 \cdot x_2 + \sqrt{3} \cdot x_1 \cdot x_2^2 + x_2^3 + \sqrt{3} \cdot x_1^2 + \sqrt{6} \cdot x_1 \cdot x_2 + \sqrt{3} \cdot x_2^2 + \sqrt{3} \cdot x_1 + \sqrt{3} \cdot x_2 + \sqrt{1})$$
$$(y_1^3 + \sqrt{3} \cdot y_1^2 \cdot y_2 + \sqrt{3} \cdot y_1 \cdot y_2^2 + y_2^3 + \sqrt{3} \cdot y_1^2 + \sqrt{6} \cdot y_1 \cdot y_2 + \sqrt{3} \cdot y_2^2 + \sqrt{3} \cdot y_1 + \sqrt{3} \cdot y_1 + \sqrt{1})$$

$$= \Psi(x) \cdot \Psi(y)$$

b. (2 pts) What did we call the function ψ in class if we remove all coefficients?

b) full rational Variaty of increasing degrees

c. (3 pts) How many multiplication operations do we save by using $K(x,y)$ versus $\psi(x) \cdot \psi(y)$?

c) ב $\Psi(x) \cdot \Psi(y)$ ישנם 10 מימדים וסה"כ 46 הכפלות (כולל מכפלות פנימיות),

וב $K(x,y)$ יש 4 הכפלות - $(x_1 \cdot y_1 + x_2 \cdot y_2 + 1)^3$. לכן בשימוש ב Kernel חוסכים 42 הכפלות.

4. Lagrange multipliers (15 pts)

Let $f(x,y) = 2x - y$. Find the minimum and the maximum points for f under the constraint

$$g(x,y) = \frac{x^2}{4} + y^2 = 1.$$

Let $f(x,y) = 2x - y$. Find the minimum and the maximum points for f under the constraint $g(x,y) = \frac{x^2}{4} + y^2 = 1$ $\left(\frac{x^2}{4} + y^2 - 1 = 0\right)$

4) תחילה נבנה את הלגרנז'יאן:

$$L(x,y,\lambda) = f(x,y) - \lambda(g(x,y) - 1) \longrightarrow L(x,y,\lambda) = 2x - y - \lambda\left(\frac{x^2}{4} + y^2 - 1\right)$$

כעת נחשב:

$$\nabla L(x,y,\lambda) = 0$$

$$\frac{\partial}{\partial x} L(x,y,\lambda) = 2 - \lambda\left(\frac{x}{2}\right) = 0$$

$$2 = \frac{\lambda x}{2}$$

$$\frac{4}{x} = \lambda$$

$$\frac{\partial}{\partial y} L(x,y,\lambda) = -1 - 2\lambda y = 0$$

$$-1 = 2\lambda y$$

$$-\frac{1}{2} = \lambda y$$

$$-\frac{1}{2} = \frac{4y}{x}$$

$$x = -8y$$

$$\boxed{x = \frac{-8}{\sqrt{17}}} \qquad \boxed{x = \frac{8}{\sqrt{17}}}$$

$$\frac{\partial}{\partial \lambda} L(x,y,\lambda) = -\frac{x^2}{4} - y^2 + 1 = 0$$

$$1 = \frac{x^2}{4} + y^2$$

$$1 = \frac{(-8y)^2}{4} + y^2$$

$$1 = 16y^2 + y^2$$

$$\frac{1}{17} = y^2$$

$$\boxed{\pm\frac{1}{\sqrt{17}} = y}$$

נציב את הנקודות בפונקציה f:

$$f\left(\frac{-8}{\sqrt{17}}, \frac{1}{\sqrt{17}}\right) = 2 \cdot \frac{-8}{\sqrt{17}} - \frac{1}{\sqrt{17}} = \frac{-17}{\sqrt{17}} = -\sqrt{17} = \min$$

$$f\left(\frac{8}{\sqrt{17}}, \frac{-1}{\sqrt{17}}\right) = 2 \cdot \frac{8}{\sqrt{17}} + \frac{1}{\sqrt{17}} = \frac{17}{\sqrt{17}} = \sqrt{17} = \max$$

$$\Longrightarrow$$

$$\left(\frac{-8}{\sqrt{17}}, \frac{1}{\sqrt{17}}\right) = \min$$

$$\left(\frac{8}{\sqrt{17}}, \frac{-1}{\sqrt{17}}\right) = \max$$