תיאורטי

חלק 1

ראשית נכתוב את התנאי $\forall i \quad y_i(<w,x_i>+b) \geq 1$

נכתוב זה כ $AV \leq d$:

$$y_i(<w,x_i>+b) = y_i x_i^T w + y_i b = y_i (x_i^T | 1) \binom{\vec{w}}{b} \geq 1$$

נסמן: $A = -\begin{bmatrix} y_1 \vec{x}_1^T & | & y_1 \\ & \vdots & \\ y_m \vec{x}_m^T & | & y_m \end{bmatrix}$, $\vec{V} = \begin{pmatrix} w_1 \\ \vdots \\ w_m \\ b \end{pmatrix}$, $\vec{d} = \begin{pmatrix} -1 \\ \vdots \\ -1 \end{pmatrix}$

ונקבל את המבוקש: $A\vec{V} \leq \vec{d}$

נצ"ב בביטוי שנרצה למזער:

$$\frac{1}{2}(w^T | b) Q \binom{w}{b} + a^T \binom{w}{b} = w^T I w$$

נצ"ב $Q = 2\begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix}$ ונקבל:

$$(w^T | b) \binom{I_m w}{0} + a^T \binom{w}{b} = w^T I w + a^T \binom{w}{b}$$

נבחר $a = 0$ ונקבל את המבוקש.

גיליון ב

1

א)

כלומר ההסתברות שדוגמא תהיה ב- $\mathcal{X}'$ גדולה מ- ε.
S_m מורכב מ- m דוגמאות כאלו. נרצה את ההסתברות שאף אחת מהן לא ב- $\mathcal{X}'$.

$$P_D(S_m \cap \mathcal{X}' = \emptyset) = P_D^m(X \notin \mathcal{X}) \le (1-\varepsilon)^m \le e^{-m\varepsilon}$$

ב) ידוע כי לכל דוגמאות x בקבוצות המקיימות S_m הפונקציה $\hat{f}$ הנתונה איננה טועה. נסמן ב $Z_{\hat{f}_s}$ את קבוצת הדוגמאות שעליהן הפונקציה טועה ($S_m \cap Z_{\hat{f}_s} = \emptyset$).

לכן ניתן לתרגם את ההסתברות מסעיף א, להסתברות שיבחרו איברים ב- $Z_{\hat{f}_s}$.

לפי סעיף א: $P_D(S_m \cap Z_{\hat{f}_s} = \emptyset) < e^{-m\varepsilon}$

כלומר $P(L_{D,f}(\hat{f}_s) > \varepsilon) = P_D(S_m \cap Z_{\hat{f}_s} = \emptyset) < e^{-m\varepsilon}$

ג) לכל היותר קיימות $|H|$ פונקציות "רעות", ולכן לפי union bound ניתן לחסום את הביטוי המבוקש בצורה הבאה:

$$P_D(L_{D,f}(A(S)) > \varepsilon) < |H| e^{-m\varepsilon}$$

נגדיר $m \geq \frac{\ln(|H|)+\ln(\frac{1}{\delta})}{\varepsilon}$ ונקבל:

$$P_D(L_{D,f}(A(S))>\varepsilon) < |H|\frac{1}{|H|}\cdot\delta = \delta$$

אנו מקבלים את ההסתברות המשלימה:

$$\boxed{P_D(L_{D,f}(A(S))>\varepsilon) > 1-\delta}$$

(2)

נבחן את האלגוריתם הבא:

נבחר את $x^* \in S_m$ עבורו $|x^*|$ מינימלי מבין הדגימות אשר עבורם $y(x)=1$. נסמן עבור איבר זה $|x^*|=a$.

נבחר את הפונקציה h_a מתוך מחלקת הפונקציות.

נניח את המקרה הריאליזבילי, יהי h_b הפונקציה הנכונה, לתיוג זה.

$a \geq b$ מכיוון שאם לא אז יש עבור x ידוע ש- $h_b(x^*)=0$ למרות ש- $y(x^*)=1$, כלומר h_b שגויה, סתירה.

נשים לב שהמקרה היחיד בו הפונקציה h_a שגויה הוא עבור $a\le|x|<b$, שכן $h_b(x)=y(x)=1$ אבל $h_a=0$.

נסמן את ההסתברות שיפול איבר כזה ב- ε.

$$P(L(A(S_m))>0)=P(a<|x|<b \text{ עבור } x \text{ ב-} S_m \text{ לכל})=$$

$$=P^m(a<|x|<b)=(1-\varepsilon)^m\leq e^{-m\varepsilon}$$

עבור $m\geq\frac{\ln(1/\delta)}{\varepsilon}$ נקבל

$$P(L(A(S_m))>0)<e^{-m\varepsilon}<\delta$$

כלומר $P(L(A(S_m))=0)>1-\delta$

3 ניתן שיש התכנסות במידה שווה, לכן קיימת קבוצת מדגם S בגודל $m_S\geq m^{uc}(\frac{\varepsilon}{2},\delta)$ כך ש-S הוא $\frac{\varepsilon}{2}$ מייצג בהסתברות $1-\delta$.

כלומר $|L_S(h)-L_D(h)|<\frac{\varepsilon}{2} \quad \forall h\in H$

$$\Downarrow$$

$$\begin{cases} L_S(h)<L_D(h)+\frac{1}{2}\varepsilon & \text{א} \\ L_D(h)<L_S(h)+\frac{1}{2}\varepsilon & \text{ב} \end{cases}$$

יהי $h^*\in H$ כך ש- $L_S(h^*)=\min\{L_S(h)\mid h\in H\}$

ויהי $h_{opt}\in H$ כך ש- $L_D(h_{opt})=\min\{L_D(h)\mid h\in H\}$

$$L_D(h^*)\leq L_S(h^*)+\frac{\varepsilon}{2}\leq L_S(h_{opt})+\frac{\varepsilon}{2}\leq L_D(h_{opt})+\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=L_D(h_{opt})+\varepsilon$$

מכיוון שקיבלנו את ההגדרה h Agnostic-PAC:

$$P_D(L_D(h^*)<L(h_{opt})+\varepsilon)>1-\delta$$

חלק 3

(1) ישנם 2^n אפשרויות לבחור את I ולכן $|H| = 2^n$.

$$Vcdim \le \log_2|H| = n$$

נראה שקיימת קבוצה בגודל n ש-H מנפצת:

נסמן $e_1, \dots, e_n$ וקטורי בסיס ב-$\mathbb{R}^n$. יהיו $y_1, \dots y_n$ תוויות המתאימות לוקטורי הבסיס.

כעת נקח את הקבוצה I כך: $I = \{ i \mid 1 \le i \le n, y_i = 1 \}$.

$$h_I(e_i) = \left(\sum_{j \in I} x_j \right) \bmod 2 = y_i$$

כלומר מצאנו קבוצה עבורה H מנפצת ולכן $Vcdim\, H = n$.

(2) נסמן $Vcdim(H_1) = n$, יהי $A = \{a_1, \dots, a_n\}$ קבוצה עליה H_1 מנפצת

יהי $h_j \in H_1$ פונקציה אשר עבורה $h(a_i) = y_i$ כאשר y_i היא התווית המתאימה ל-a_i.

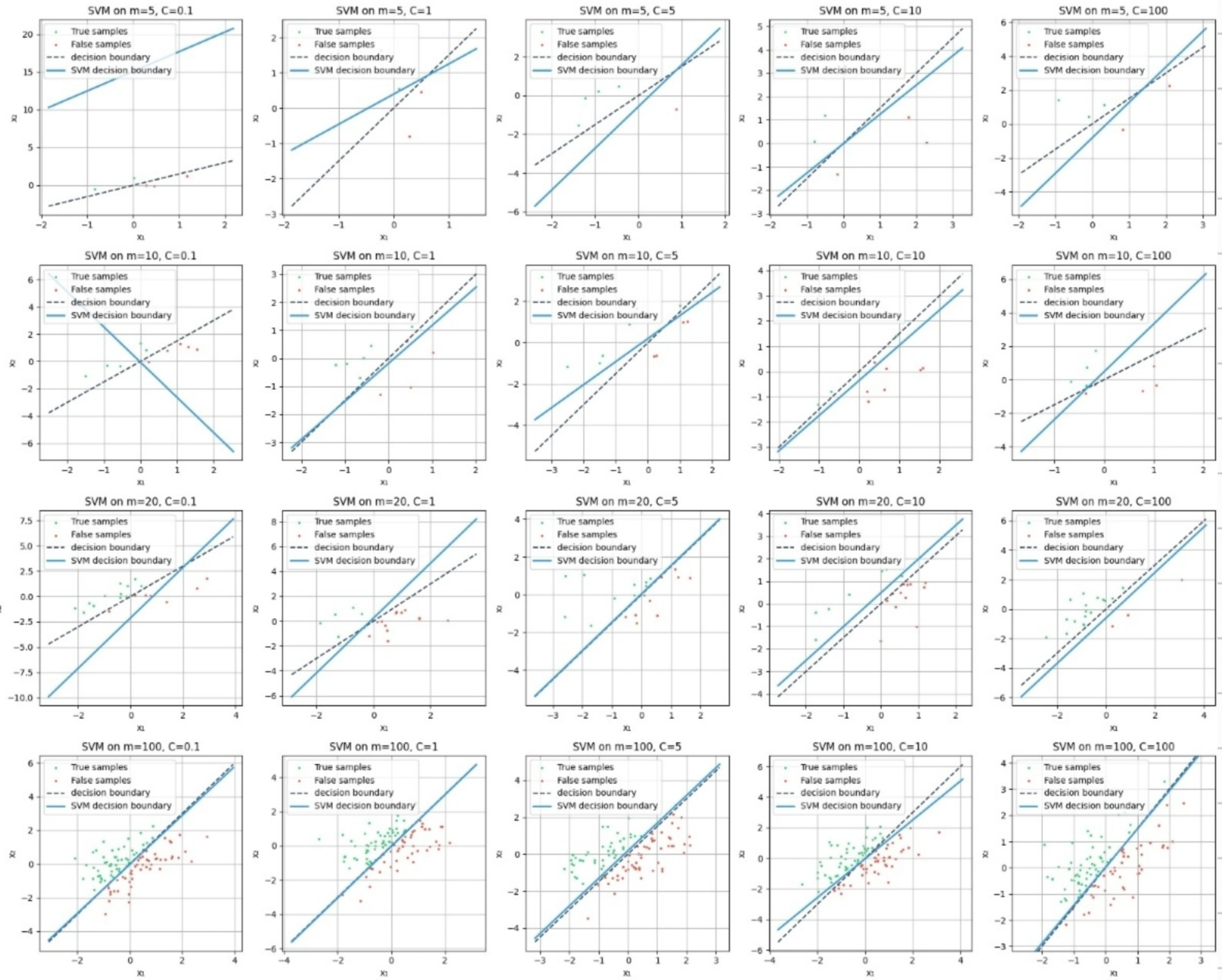
$H_1 \subseteq H_2$ ולכן הפונקציה h נמצאת גם ב-H_2. לכן H_2 גם מנפצת את A, כלומר $Vcdim(H_2) \ge n$.

כעת $H_1 \subseteq H_2 \Rightarrow |H_1| \le |H_2|$ ולכן מתקיים כמובן גם כך:

$$n = Vcdim(H_1) < \log_2|H_1| \le Vcdim(H_2) \le \log_2|H_2|$$

גלעד מנשה

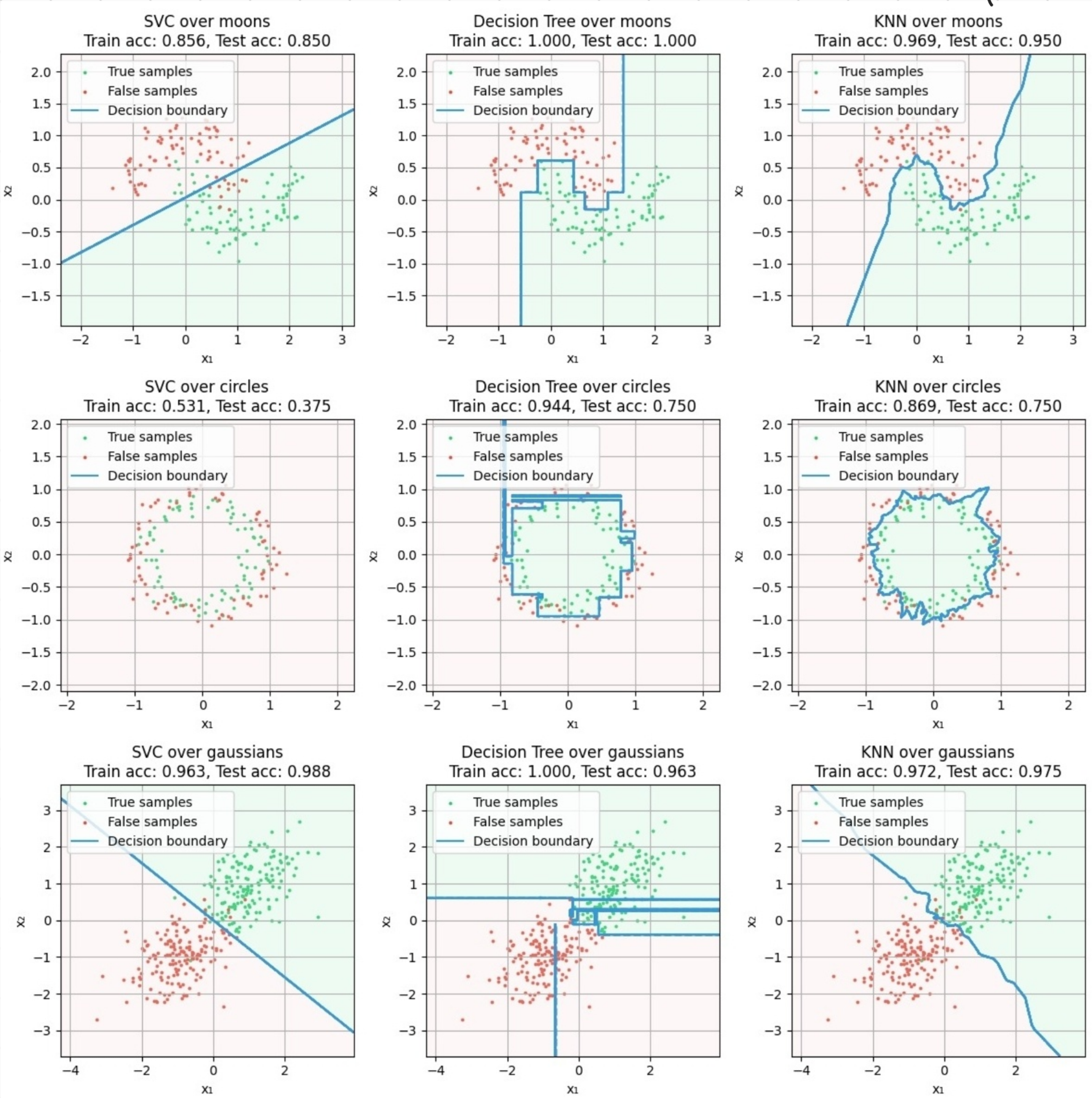
1) א)



ב) לפי הגרפים נראה כי יש נקודות מסויימות עבורן הפרמטר C משנה את הביצוע של המודל. נראה שעבור C נמוך ה-margin גדול והמודל עושה טעויות משמעותיות. עבור C גדול השגיאה קטנה אך נראה שההשפעה פחות ניכרת לפי הדוגמאות.

ג) ככל שהפרמטר m גדל, כך המודל משיג תוצאות טובות יותר. לפי הא"י שוויון שראינו בכיתה עבור $m(\epsilon,\delta)$. ככל שמספר הדוגמאות גדל כך ניתן להגיע לתת בעיקר שגיאה קטנה יותר. התוצאה הזו באה לידי ביטוי בגרפים.

2 א)



ד) SVM - המודל הזה מבדיל את סוג הנתונים ע"י קו מפריד.
לכן המודל הזה יהיה בעל ביצועים שגויים כאשר בנקודות המדגם יש
קישור בין הלייבלים
ולכן ראינו שעבור פיזור מעגלי SVM בעל דיוק נמוך.

עץ החלטה - המודל מוגדר על סיווג האיזורים לפי כמות הלייבלים ביניהן, לכן מודל זה יעניק לנו כמשן הדגימות בצורה מדויקת זהה מקורבות.

ממאן - המודל מסווג לפי שיטת שכנים קרובים, לכן כמו בעץ החלטה, המודל יהיה מיטבי כאשר יש איזורי הפרדה ברורים בין הלייבלים.

ג) מודל מחפש את הקו אשר יפריד באופן מיטבי את הלייבלים ולכן גבול ההחלטה יהיה קו מפריד בין איזורים.

עץ ההחלטה מחלק אנך המרחק לחלונים וכל פעם קובע את הלייבל לפי הרוב. לכן נקבל שגבול ההחלטה יהיה גבולות קווים ישרים מאונכים לצירים.

ממאן קובע את האיזורים לפי השכנים הקרובים ולכן גבול ההחלטה יעבור בין נקודות בהם יש שינוי מובהק בלייבלים של נקודות המדגם.