Short HW3 – SVM, Optimization, and PAC learning

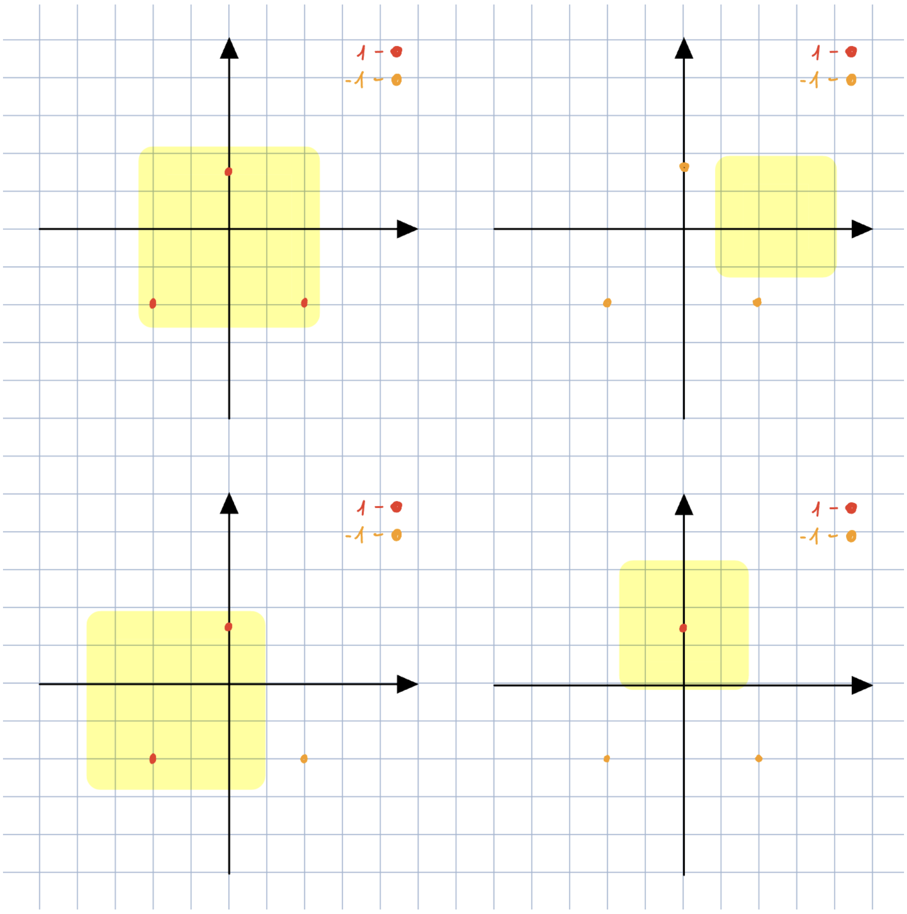
1. טענה: $VCdim(\mathcal{H}_{sqr}) = 3$.

על מנת להוכיח את טענה זו, עלינו להראות:

1. קיים סט של 3 דוגמאות אשר ניתן לניתוץ, ומכאן נסיק כי $VCdim(\mathcal{H}_{sqr}) \geqslant 3$.
2. לא קיים סט של 4 דוגמאות אשר ניתן לניתוץ, ומכאן נסיק כי $VCdim(\mathcal{H}_{sqr}) < 4$.

משתי הטענות יחד נוכל להראות כי $VCdim(\mathcal{H}_{sqr}) = 3$.

תחילה נוכיח את 1 - נראה סט של 3 דוגמאות אשר ניתן לניתוץ, נסדר את הדוגמאות כמשולש שווה צלעות ונקבל:



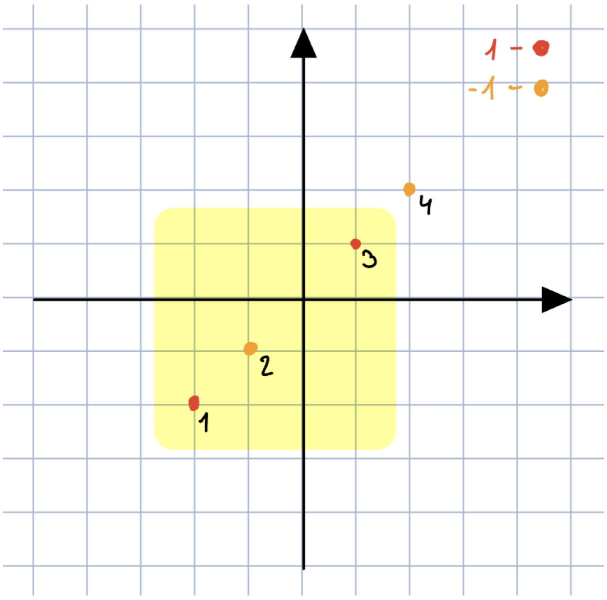
בדוגמה זו לכל תיוג אפשרי של 3 הדוגמאות קיימת היפותזה מתאימה אשר מתייגת את שלושתן עם שגיאה של 0, כלומר סט דוגמאות זה ניתן לניתוץ על ידי מחלק ההיפותזות הנתונה. הריבוע הצהוב מייצג את האיזור שבו ההיפותה מתייגת את הדוגמה כ-1 ומחוץ לריבוע הצהוב כ--1. (הנחתי כי שינוי סדר הדוגמאות בסידורים בעלי מספר דוגמאות חיוביות ושליליות זהה לא משפיע על היתכנות קיום היפותזה מתאימה מפני שניתן להעביר את הריבוע הצהוב בהתאם לסידור הדוגמאות - כלומר שמשיקולי סימטריה סידור הדוגמאות החיוביות והשליליות לא משפיע על קיום התיוג), כלומר קיים סט של 3 דוגמאות שניתן לניתוץ, לכן לפי משפט מהתרגול מתקיים כי $VCdim(\mathcal{H}_{sqr}) \geqslant 3$.

כעת נוכיח את 2 - נראה כי לכל סט של 4 דוגמאות לא קיימת מחלקת היפותזות אשר מנתצת את הסט:

נחלק את ההוכחה למקרים לפי מיקום הנקודות במרחב:

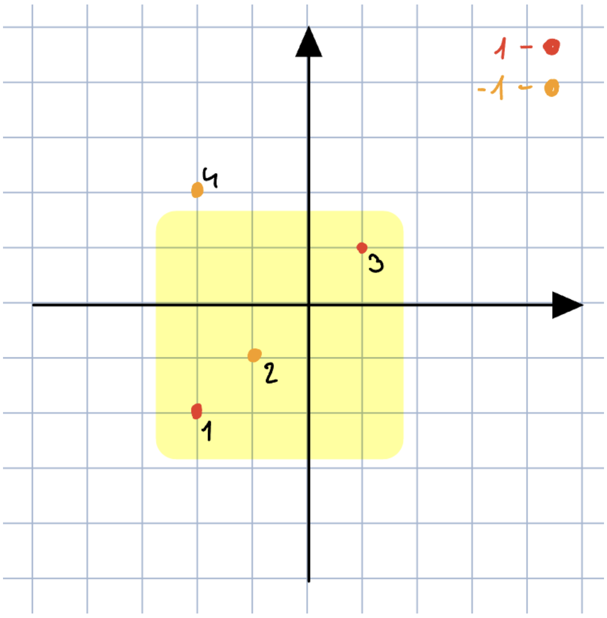
- אם כל הדוגמאות נמצאות על ישר, נסמן אותן לפי ערכי ה-x שלהן (כלומר 1 תהיה בעלת ערך ה-x

הקטן ביותר ו-4 בעלת ערך ה-x הגדול ביותר). עבור תיוג של 1,3 כחיוביים ו2,4 כשליליים נקבל:



כלומר עבור כל ריבוע שנגדיר הוא יצטרך לכלול את דוגמאות 1,2,3 כלומר תהיה שגיאה כלשהי ולא נוכל לתייג את ארבעת הדוגמאות עם שגיאה של 0.

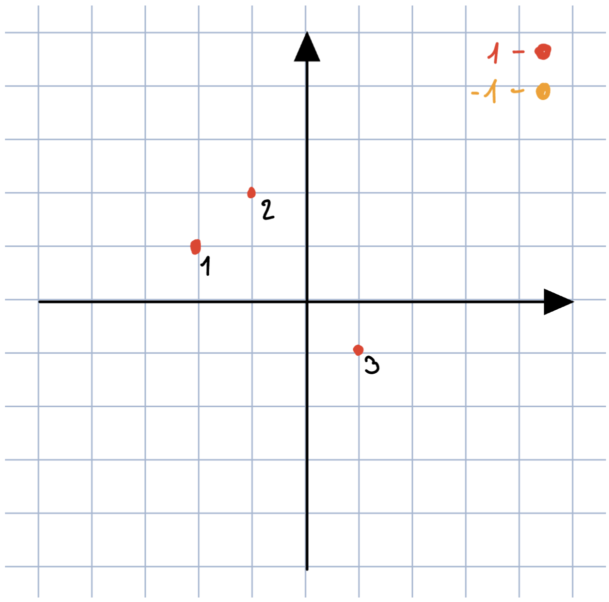
- עבור שלוש נקודות המסודרות על ישר ונקודה רביעית כלשהי נקבל:



ומאותם נימוקים כמו המקרה הקודם נקבל כי דוגמה 2 תמצא בין 1-3 בציר ה-x ולכן תהיה כלולה בכל היפותזה אפשרית שנבחר, כלומר השגיאה של ההיפותזה לא תהיה 0.

- אם שני המקרים הקודמים לא מתקיימים, אז נטען כי קיימות שתי דוגמאות שדרכן נוכל להעביר ישר כלשהו, כאשר שתי הנקודות הנותרות לא יהיו חלק מהמלבן שיווצר בין הקואורדינטות של שתי הנקודות הראשונות

נסמן את ארבעת הנקודות במרחב כך: $1:(x_1,y_1),\ 2:(x_2,y_2),\ 3:(x_3,y_3),\ 4:(x_4,y_4),$



מתקיים בדוגמה זו $x_1 < x_2 < x_3$ ובנוסף $y_3 < y_1 < y_2$, כלומר הבטחנו כי דוגמה 3 לא נמצאת במלבן
הנוצר על ידי הדוגמאות $1, 2$.
נבחן את כל המקרים האפשריים עבור הדוגמה 4, נניח כי נקודה זו לא נמצאת בתוך מלבן הנוצר על ידי שתי נקודות כלשהן, כלומר לא ייתכן כי $x_4 < x_1$ וגם כי $y_4 > y_2$ כי במצב זה המלבן שיווצר על ידי הדוגמאות 3,4 יכלול את הדוגמאות 1,2 ובאותו אופן לא ייתכן כי $x_4 < x_1$ וגם $y_4 < y_2$ כי אז המלבן שיווצר על ידי דוגמאות 2,4 יכיל את דוגמה 1.
מצירוף שתי המקרים שתוארו מעלה נקבל כי לא ייתכן כי $x_4 < x_1$ כלומר בוודאות מתקיים $x_4 > x_1$ אך כן ייתכן מקרה שבו נבחר תיוג שכזה, ובמקרה זה הוא יהיה בעל שגיאה גדולה מ-0 עבור תיוג של דוגמה 4 כ--1.
מכאן נסיק כי בהכרח קיימות זוג נקודות שעל פיהן מוגדר הריבוע שמוגדר על ידי ההיפתזה וריבוע זה בהכרח יכיל נקודה שלישית כלשהי.
עבור כל תיוג של 2 הדוגמאות 1,2 כ-1 , כל ריבוע שיכיל את שתיהן יכיל גם נקודה נוספת, עבור נקודה כזו שתתיוג כ--1 נקבל כי בהכרח ישנה שגיאה כלשהי ולא נקבל היפותזה בעלת שגיאה 0.
כיסינו את כל המקרים האפשריים לחלוקת 4 הנקודות במרחב וקיבלנו כי לא קיימת היפותזה ממחלקת ההפיתזות אשר מנתצת את סט הדוגמאות, כלומר $VCdim(\mathcal{H}_{sqr}) < 4$.
משתי הטענות שהוכחו יחד נקבל כי $VCdim(\mathcal{H}_{sqr}) = 3$.

2. נתונים שני kernels:

$$K_1(u, v) = \langle \phi_1(u), \phi_1(v) \rangle$$
$$K_2(u, v) = \langle \phi_2(u), \phi_2(v) \rangle$$

כאשר ϕ_1, ϕ_2 מוגדרות כך:

$$\phi_1 : X \to \mathbb{R}^{n_1}, n_1 \in \mathbb{N}$$
$$\phi_2 : X \to \mathbb{R}^{n_2}, n_2 \in \mathbb{N}$$

נראה כי ה-kernel החדש שהוגדר:

$$K_3 = 2K_1(u, v) + 3K_2(u, v)$$

גם כן kernel חוקי, כלומר שקיימת $\phi_3 : X \to \mathbb{R}^{n_3},\ n_3 \in \mathbb{N}$ כך שמתקיים:

$$K_3(u,v) = \langle \phi_3(u), \phi_3(v) \rangle$$

נגדיר את ϕ_3 באופן הבא:

$$\phi_3(u) = \begin{bmatrix} \sqrt{2} \cdot \phi_1(u) \\ \sqrt{3} \cdot \phi_2(u) \end{bmatrix}$$

$$\implies \langle \phi_3(u), \phi_3(v) \rangle = \langle \begin{bmatrix} \sqrt{2} \cdot \phi_1(u) \\ \sqrt{3} \cdot \phi_2(u) \end{bmatrix}, \begin{bmatrix} \sqrt{2} \cdot \phi_1(v) \\ \sqrt{3} \cdot \phi_2(v) \end{bmatrix} \rangle$$

$$= \begin{bmatrix} \sqrt{2} \cdot \phi_1(u) \\ \sqrt{3} \cdot \phi_2(u) \end{bmatrix}^T \cdot \begin{bmatrix} \sqrt{2} \cdot \phi_1(v) \\ \sqrt{3} \cdot \phi_2(v) \end{bmatrix} = \begin{bmatrix} \sqrt{2} \cdot \phi_1(u) & \sqrt{3} \cdot \phi_2(u) \end{bmatrix}^T \cdot \begin{bmatrix} \sqrt{2} \cdot \phi_1(v) \\ \sqrt{3} \cdot \phi_2(v) \end{bmatrix}$$

$$= 2 \cdot \phi_1^T(u) \cdot \phi_1(v) + 3 \cdot \phi_2^T(u) \cdot \phi_2(v) = 2 \cdot \langle \phi_1(u), \phi_1(v) \rangle + 3 \cdot \langle \phi_2(u), \phi_3(v) \rangle$$

$$= 2K_1(u,v) + 3K_2(u,v) = K_3$$

לפי הגדרת ϕ_3 נבחין כי $n_3 = n_1 + n_2$.

כלומר קיבלנו כי קיימת $\phi_3 : X \to \mathbb{R}^{n_3},\ n_3 \in \mathbb{N}$ כך שמתקיים $K_3(u,v) = \langle \phi_3(u), \phi_3(v) \rangle$, כלומר K_3 הוא kernel חוקי.

.3

1. יהיו פונקציות $f, g : C \to \mathbb{R}$ פונקציות קמורות מעל קבוצה קמורה C.

נוכיח תחילה לפי הגדרה כי $q(z) \triangleq max\{f(z), g(z)\}$ היא פונקציה קמורה:

יהיו $z_1, z_2 \in C,\ t \in [0,1]$, נרצה להראות כי מתקיים:

$$q(t \cdot z_1 + (1-t) \cdot z_2) \leqslant t \cdot q(z_1) + (1-t) \cdot q(z_2)$$

תחילה, נתון כי f, g הן קמורות ולכן מתקיים:

$$f(t \cdot z_1 + (1-t) \cdot z_2) \leqslant t \cdot f(z_1) + (1-t) \cdot f(z_2)$$
$$g(t \cdot z_1 + (1-t) \cdot z_2) \leqslant t \cdot g(z_1) + (1-t) \cdot g(z_2)$$

לפי הגדרת $q(z)$:

$$q(t \cdot z_1 + (1-t) \cdot z_2) \underbrace{=}_{\text{הגדרת } q} max\{f(t \cdot z_1 + (1-t) \cdot z_2),\ g(t \cdot z_1 + (1-t) \cdot z_2)\}$$
$$\underbrace{\leqslant}_{\text{קמורות } f,g} max\{t \cdot f(z_1) + (1-t) \cdot f(z_2),\ g \cdot g(z_1) + (1-t) \cdot g(z_2)\}$$
$$\underbrace{=}_{(1-t),t \geqslant 0} t \cdot max\{f(z_1), g(z_1)\} + (1-t) \cdot max\{f(z_2), g(z_2)\}$$
$$\underbrace{=}_{\text{הגדרת } q} t \cdot q(z_1) + (1-t) \cdot q(z_2)$$

כלומר הוכחנו את הנדרש, ולכן $q(z)$ פונקציה קמורה.

2. לפי הלמה שנלמדה בתרגול 7, כל פונקציה לינארית היא קמורה, בפרט $1 - y_i w^T x_i$ שהיא פונקציה לינארית היא גם קמורה, גם הפונקציה $f = 0$ היא פונקציה קבועה כלומר לינארית ולכן גם היא קמורה.

בסעיף 3.1 הראנו כי פונקציית מקסימום בין שתי פונקציות קמורות היא קמורה גם כן ולכן $max\left\{0, 1 - y_i w^T x_i\right\}$ היא קמורה כנדרש.

3. נראה כי בעיית ה-soft SVM היא קמורה, עבור:

$$\underset{w \in R^d}{\arg\min} \frac{1}{m} \sum_{i=1}^{m} max\left\{0, 1 - y_i w^T x_i\right\} + \lambda ||w||_2^2$$

תחילה, בתרגול ראינו כי $||w||_2^2$ היא פונקציה קמורה (כסכום של פונקציות ריבועיות), לפי הלמה הנתונה בשאלה לכל $\mathbb{R} \ni \alpha \geqslant 0$ מתקיים כי אם $f(z)$ קמורה אז גם $\alpha f(z)$ קמורה, ובפרט עבור $f = ||w||_2^2, \ \lambda \geqslant 0$ נקבל כי $\lambda ||w||_2^2$ היא פונקציה קמורה.

כעת, לפי סעיף 3.2 הראנו כי $max\left\{0, 1 - y_i w^T x_i\right\}$ היא פונקציה קמורה גם כן, בתרגול ראינו כי גם סכום של פונקציות קמורות הוא קמור ולכן $\sum_{i=1}^{m} max\left\{0, 1 - y_i w^T x_i\right\}$ היא פונקציה קמורה, בתרגול גם הוצג כי ממוצע של פונקציות קמורות גם כן קמור ולכן $\frac{1}{m} \sum_{i=1}^{m} max\left\{0, 1 - y_i w^T x_i\right\}$ היא פונקציה קמורה.

נשתמש בתכונה של חיבור שתי פונקציות קמורות בשנית ונקבל כי הביטוי $\frac{1}{m} \sum_{i=1}^{m} max\left\{0, 1 - y_i w^T x_i\right\} + \lambda ||w||_2^2$ גם כן קמור, ולבסוף איבר מינימלי מבין קבוצה של איברים קמורים הוא בפרט קמור.

סך הכל הראנו כי הביטוי המייצג את בעיית ה-soft SVM הוא קמור.