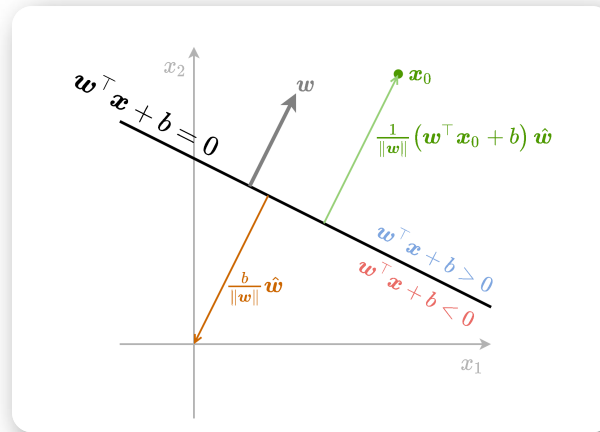


תרגול 6 - SVM ופונקציות גרעין (Kernels)

הערה: בפרק זה נעסוק בסיווג בינארי ונסמן את התוויות של שתי המחלקות ב 1 ו -1 .

תזכורת - גאומטריה של המישור



- מישור מתואר באופן הבא $w^T x + b = 0$.
- המישור מחלק את המרחב ומגדיר צד חיובי וצד שלילי.
- w הינו הנורמל למישור.
- המרחק של המישור מהראשית הינו $\frac{b}{\|w\|}$.
- המרחק של נקודה x_0 מהמישור הינה $\frac{1}{\|w\|} (w^T x_0 + b)$.
- w ו b מוגדרים עד כדי קבוע.

מסוג לינארי

מסוג לינארי הוא מסוג מהצורה של

$$h(\mathbf{x}) = \text{sign}(\mathbf{w}^\top \mathbf{x} + b) = \begin{cases} 1 & \mathbf{w}^\top \mathbf{x} + b > 0 \\ -1 & \text{else} \end{cases}$$

עם w ו- b כל שהם.

• המסוג מחלק את המרחב לשני צידיו של המישור:

$$\mathbf{w}^\top \mathbf{x} + b = 0 \quad \circ$$

• מישור זה מכונה מישור הפרדה.

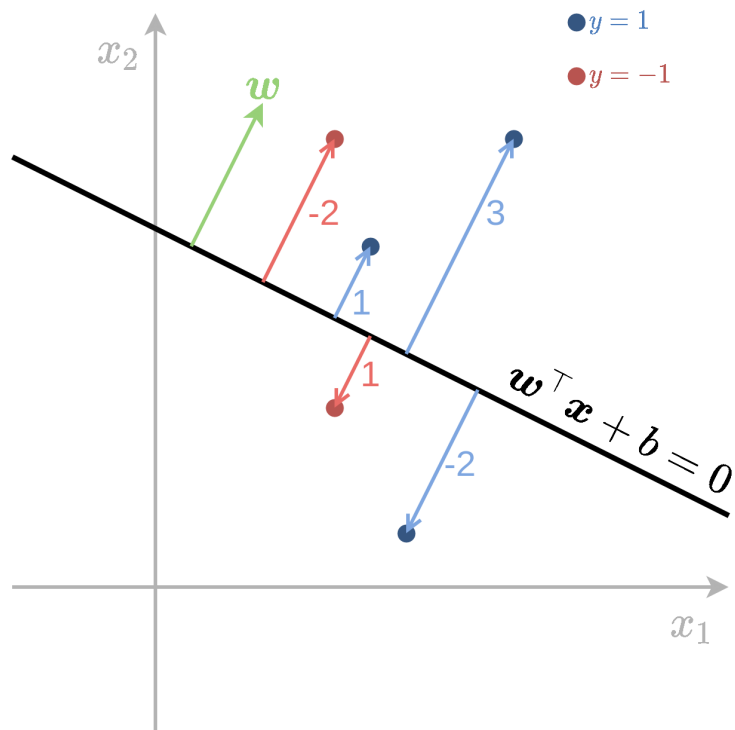
Signed distance (מרחק מסומן)

נסתכל על בעיית סיווג בינארית עם תוויות $y = \pm 1$.

- נגדיר את ה- signed distance של דגימה כל שהיא ממשטח ההפרדה באופן הבא:

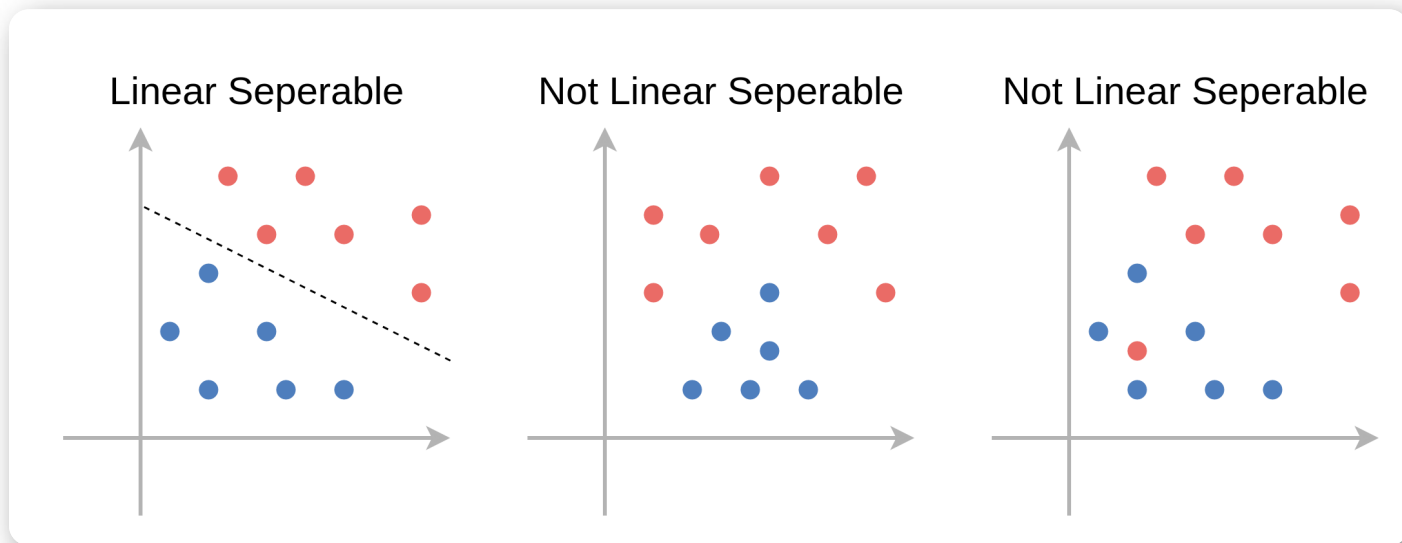
$$d = \frac{1}{\|w\|} (w^\top x + b)y$$

- זהו המרחק של נקודה מהמישור.
- כאשר y תואם למיקום הנקודה x ביחס למישור, הערך חיובי.



פרידות לינארית (linear separability)

- בבעיות סיווג בינארי, נאמר על המדגם שהוא פריד לינארית אם קיים מסווג לינארי אשר מסווג את המדגם בצורה מושלמת (בלי טעויות סיווג).
- כאשר המדגם פריד לינארית יהיו יותר ממסווג לינארי אחד אשר יכול לסווג בצורה מושלמת את המדגם.



(Support Vector Machine (SVM

- אלגוריתם דיסקרימינטיבי לסיווג בינארי אשר מחפש מסווג לינארי אשר יסווג בצורה טובה את המדגם.
- אלגוריתם זה יבחר את המסווג בעל השוליים המירביות מכל המסווגים שמצליחים לסווג את המדגם.

- ניתן לשימוש אך ורק כאשר המדגם פריד לינארית.
- נחפש את המסווג הלינארי אשר בעבורו ה signed distance המינימאלי על ה train set הוא המקסימאלי:

$$w^*, b^* = \arg \max_{w, b} \min_i \left\{ \frac{1}{\|w\|} (w^\top x^{(i)} + b) y^{(i)} \right\}$$

- בעיה זו שקולה לבעיית האופטימיזציה הבאה:

$$w^*, b^* = \arg \min_{w, b} \frac{1}{2} \|w\|^2$$

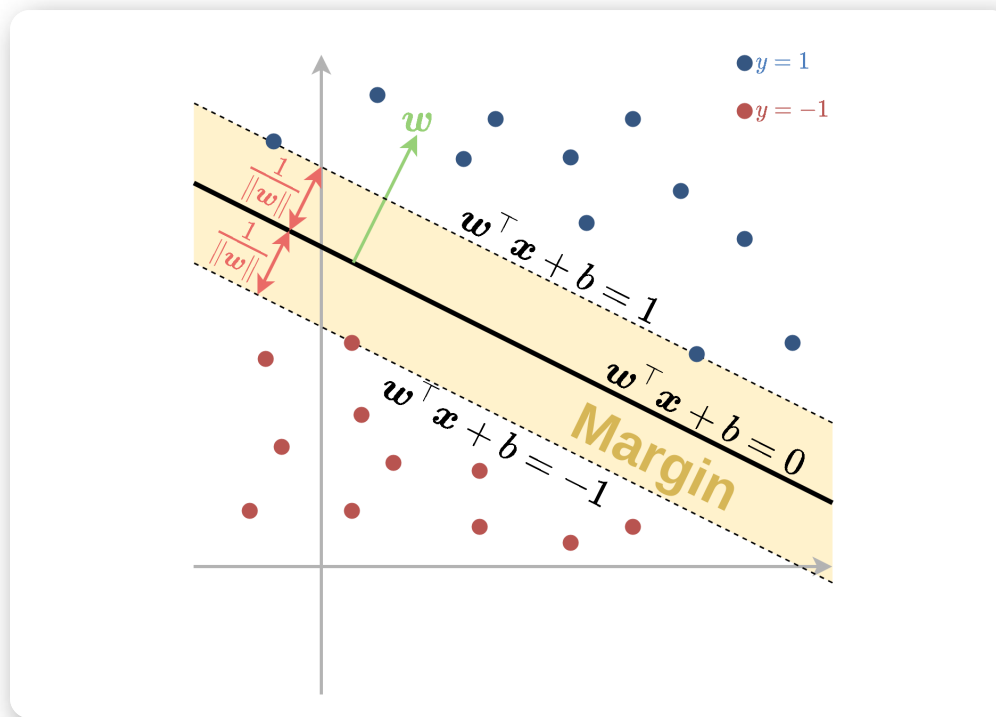
s.t. $y^{(i)} (w^\top x^{(i)} + b) \geq 1 \quad \forall i$

- בעיית אופטימיזציה זו מכונה הבעיה הפרימאלית.

Margin - שוליים

- בכדי להבין את המשמעות של בעיית האופטימיזציה שקיבלנו נגדיר את המושג שוליים (margin) של המסווג.
- האיזור של ה margin מוגדר כאיזור סביב משטח ההפרדה אשר נמצא בתחום:

$$1 \geq w^T x + b \geq -1$$



- רוחב איזור זה נקבע על פי הגודל של הוקטור w ושווה ל $\frac{2}{\|w\|}$

- האילוץ $y^{(i)} (w^T x^{(i)} + b) \geq 1$ אשר מופיע בבעיית האופטימיזציה הפרימאלית, דורש למעשה שכל הנקודות יסווגו נכונה **וימצאו מחוץ ל margin**.

- בעיית האופטימיזציה מחפשת את הפרמטרים של המישור בעל ה margin הגדול ביותר אשר מקיים תנאי זה.

Support Vectors

- ה- **support vectors** מוגדרים כנקודות אשר יושבות על השפה של ה **margin**, עבור פתרון של בעיית האופטימיזציה.

- נקודות אלו מקיימות $y^{(i)} (w^T x^{(i)} + b) = 1$.

- אלו הנקודות אשר ישפיעו על הפתרון של בעיית האופטימיזציה.

- לכן, הסרה או הזזה קטנה של נקודות שאינן **support vectors** לא תשנה את הפתרון.

הבעיה הדואלית

- דרך שקולה לרישום בעיית האופטימיזציה הינה על ידי הגדרת N משתני עזר נוספים $\{\alpha_i\}_{i=1}^N$.
- בעזרת משתנים אלו ניתן לרשום את בעיית האופטימיזציה באופן הבא:

$$\begin{aligned} \{\alpha_i\}^* &= \arg \max_{\{\alpha_i\}} \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y^{(i)} y^{(j)} \alpha_i \alpha_j \mathbf{x}^{(i)\top} \mathbf{x}^{(j)} \\ \text{s.t. } &\alpha_i \geq 0 \quad \forall i \\ &\sum_i \alpha_i y^{(i)} = 0 \end{aligned}$$

מתוך המשתנים $\{\alpha_i\}_{i=1}^N$ ניתן לשחזר את w אופן הבא:

$$w = \sum_i \alpha_i y^{(i)} \mathbf{x}^{(i)}$$

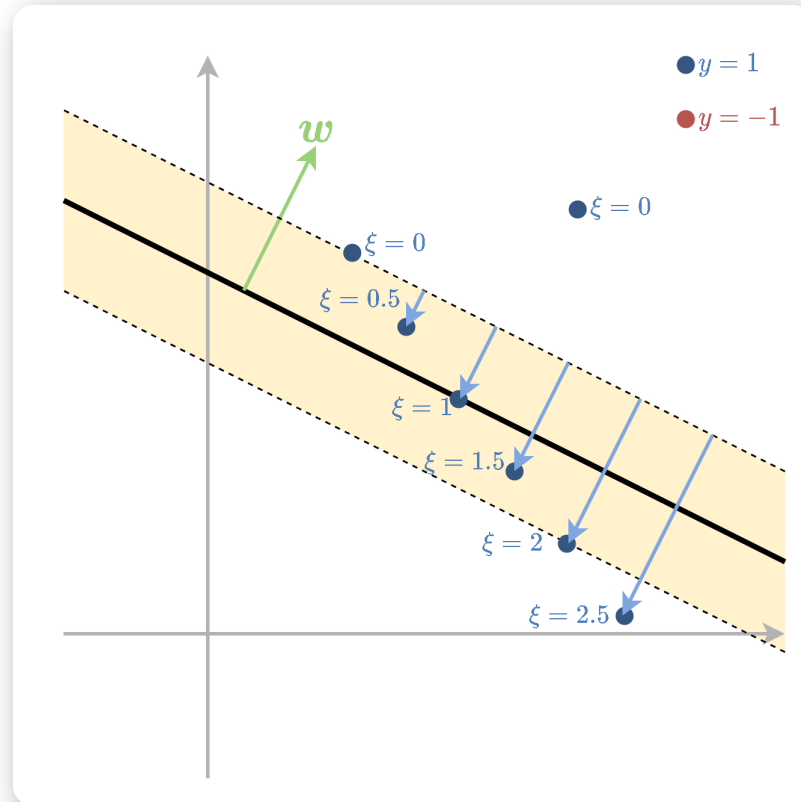
קשר בין הבעיה הדואלית והפרימאלית:

$\alpha_i = 0$	$y^{(i)} (w^\top x^{(i)} + b) > 1$	נקודות רחוקות מה margin
$\alpha_i \geq 0$	$y^{(i)} (w^\top x^{(i)} + b) = 1$	נקודות על ה margin (שהם support vectors)

- אם $\alpha_i > 0$ אז $y^{(i)} (w^\top x^{(i)} + b) = 1$ (אבל לא להיפך)
- אם $y^{(i)} (w^\top x^{(i)} + b) > 1$ אז $\alpha_i = 0$ (אבל לא להיפך)
- את b נחשב על ידי בחירת SV ונחלץ אותו מתוך $y^{(i)} (w^\top x^{(i)} + b) = 1$.

- **Soft SVM מתייחס למקרה שבו המדגם אינו פריד לינארית.**
- **במקרה זה עדיין מגדירים את ה margin בצורה דומה אך מאפשרים למשתנים להיכנס לתוך ה margin ואף לחצות אותה לצד הלא נכון של מישור ההפרדה.**
- **על כל חריגה כזו משלמים קנס ב objective שאותו מנסים למזער.**

- את החריגה של הדגימה ה i נסמן ב $\frac{1}{\|w\|} \xi_i$. לנקודות שהן בצד הנכון של המישור ומחוץ ל margin יהיה $\xi_i = 0$



- המשתנים ξ_i נקראים **slack variables**.

• בעיית האופטימיזציה הפרימאלית תהיה:

$$\begin{aligned} \mathbf{w}^*, b^*, \{\xi_i\}^* = \arg \min_{\mathbf{w}, b, \{\xi_i\}} & \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \\ \text{s.t.} & \quad y^{(i)} \left(\mathbf{w}^\top \mathbf{x}^{(i)} + b \right) \geq 1 - \xi_i \quad \forall i \\ & \quad \xi_i \geq 0 \quad \forall i \end{aligned}$$

• C הוא hyper-parameter שקובע את גודל הקנס על חריגה.

• הבעיה הדואלית הינה:

• ה **support vectors** מוגדרים להיות הנקודות שמקיימות

$$y^{(i)} (\mathbf{w}^\top \mathbf{x}^{(i)} + b) = 1 - \xi_i$$

תכונות:

•	•	•
$\alpha_i = 0$	$y^{(i)} (\mathbf{w}^\top \mathbf{x}^{(i)} + b) > 1$	נקודות שמסווגות נכון ורחוקות מה margin
$0 \leq \alpha_i \leq C$	$y^{(i)} (\mathbf{w}^\top \mathbf{x}^{(i)} + b) = 1$	נקודות על ה margin (שהם support vectors)
$\alpha_i = C$	$y^{(i)} (\mathbf{w}^\top \mathbf{x}^{(i)} + b) = 1 - \xi_i$	נקודות שחורגות מה margin (גם support vectors)

מאפיינים: תזכורת

- נוכל תמיד לחליף את וקטור המשתנים x שעליו פועל האלגוריתם בוקטור חדש $x_{\text{new}} = \Phi(x)$.
- Φ היא פונקציה אשר נבחרה מראש ונקראת פונקציית המאפיינים.

פונקציות גרעין

- החישוב של $\Phi(x)$ יכול להיות יקר.
- לעיתים, קיימת דרך לחשב בצורה יעילה את הפונקציה
 $K(x_1, x_2) = \Phi(x_1)^\top \Phi(x_2)$ אשר נקראת פונקציית גרעין
 - ישנם אפילו מקרים שבהם וקטור המאפיינים אינסופי!
- ישנם קריטריונים תחתם פונקציה מסוימת $K(x_1, x_2)$ היא פונקציית גרעין בעבור וקטור מאפיינים מסויים. בקורס לא נכנס לכך.

נציג שתי פונקציות גרעין נפוצות:

- גרעין גאוסני:

- $$K(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(-\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2}{2\sigma^2}\right)$$

- כאשר σ פרמטר שיש לקבוע.

- גרעין פולינומיאלי: $K(\mathbf{x}_1, \mathbf{x}_2) = (1 + \mathbf{x}_1^\top \mathbf{x}_2)^p$

- כאשר $p \geq 1$ פרמטר שיש לקבוע.

פונקציית המאפיינים שמתאימות לגרעינים אלו מסורבלות לכתובה ולא נציג אותן כאן.

Kernel Trick in SVM

- טריק זה מאפשר להשתמש ב SVM עם מאפיינים מבלי לחשב את Φ באופן ישיר באמצעות פונקציית גרעין.
- עבור פונקציית מאפיינים Φ עם פונקציית גרעין K הבעיה הדואלית של SVM הינה:

$$\begin{aligned} \{\alpha_i\}^* &= \arg \max_{\{\alpha_i\}} \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y^{(i)} y^{(j)} \alpha_i \alpha_j K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) \\ \text{s.t. } &\alpha_i \geq 0 \quad \forall i \\ &\sum_i \alpha_i y^{(i)} = 0 \end{aligned}$$

- בעיית זו מגדירה את המשתנים $\{\alpha_i\}$ בלי צורך לחשב את Φ באופן מפורש בשום שלב.

Kernel Trick in SVM

- הפרמטר w נתון על ידי:

$$w = \sum_i \alpha_i y^{(i)} \Phi(x^{(i)})$$

- כדי לחשב את w באופן מפורש יש לחשב את Φ .

- עם זאת, ניתן להמנע מכך

Kernel Trick in SVM

- נציב את הנוסחא של w ישירות למסווג ובכך נוכל להינמע מלחשב את Φ :

$$\begin{aligned}h(\mathbf{x}) &= \text{sign}(\mathbf{w}^\top \Phi(\mathbf{x}) + b) \\ &= \text{sign}\left(\sum_i \alpha_i y^{(i)} \Phi(\mathbf{x}^{(i)})^\top \Phi(\mathbf{x}) + b\right) \\ &= \text{sign}\left(\sum_i \alpha_i y^{(i)} K(\mathbf{x}^{(i)}, \mathbf{x}) + b\right)\end{aligned}$$

- כך בעצם נוכל להשתמש במסווג מבלי לחשב את Φ מפורשות

תרגיל 6.1 - Support Vectors 2

בשאלה זו נראה שבעבור מדגם המכיל 2 נקודות, אחת מכל מחלקה, הפתרון של בעיית hard SVM יהיה מישור הפרדה שיעבור בדיוק במרכז בין 2 הנקודות, ויהיה ניצב לוקטור המחבר את הנקודות

$$\mathbf{w} = \frac{2}{\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_2^2} (\mathbf{x}^{(1)} - \mathbf{x}^{(2)}), \quad b = -\frac{1}{2} (\mathbf{x}^{(1)} + \mathbf{x}^{(2)})^\top \mathbf{w}$$

1) הראו זאת על ידי פתרון הבעיה הדואלית.

פתרון:

הבעיה הדואלית הינה:

$$\begin{aligned} \{\alpha_i\}^* &= \arg \max_{\{\alpha_i\}} \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y^{(i)} y^{(j)} \alpha_i \alpha_j \mathbf{x}^{(i)\top} \mathbf{x}^{(j)} \\ \text{s.t. } &\alpha_i \geq 0 \quad \forall i \\ &\sum_i \alpha_i y^{(i)} = 0 \end{aligned}$$

• נניח בלי הגבלת הכלליות ש $y^{(1)} = 1$ ו $y^{(2)} = -1$.

נסתכל על האילוץ בשורה האחרונה:

$$\sum_{i=1}^2 \alpha_i y^{(i)} = 0$$

$$\Leftrightarrow \alpha_1 - \alpha_2 = 0$$

$$\Leftrightarrow \alpha_1 = \alpha_2 = \alpha$$

• מכאן שהמשתנה היחיד בבעיה הינו α .

• בעיית האופטימיזציה (ללא האילוצים) תהיה:

$$\begin{aligned}\alpha &= \arg \max_{\alpha} 2\alpha - \frac{\alpha^2}{2} \sum_{i,j} y^{(i)} y^{(j)} \mathbf{x}^{(i)\top} \mathbf{x}^{(j)} \\ &\text{s.t. } \alpha \geq 0 \\ &= \arg \max_{\alpha} 2\alpha - \frac{\alpha^2}{2} \left(\|\mathbf{x}^{(1)}\|_2^2 - \mathbf{x}^{(1)\top} \mathbf{x}^{(2)} - \mathbf{x}^{(2)\top} \mathbf{x}^{(1)} + \|\mathbf{x}^{(2)}\|_2^2 \right) \\ &\text{s.t. } \alpha \geq 0 \\ &= \arg \max_{\alpha} 2\alpha - \frac{\alpha^2}{2} \|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_2^2 \\ &\text{s.t. } \alpha \geq 0\end{aligned}$$

ניתן לפתור את הבעיה על ידי גזירה והשוואה ל-0:

$$\begin{aligned}\frac{d}{d\alpha} 2\alpha - \frac{\alpha^2}{2} \|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_2^2 &= 0 \\ \Leftrightarrow 2 &= \alpha \|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_2^2 \\ \Leftrightarrow \alpha &= \frac{2}{\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_2^2}\end{aligned}$$

את w נמצא על ידי:

$$w = \sum_i \alpha_i y^{(i)} \mathbf{x}^{(i)} = \alpha (\mathbf{x}^{(1)} - \mathbf{x}^{(2)}) = \frac{2}{\|\mathbf{x}^{(1)} - \mathbf{x}^{(2)}\|_2^2} (\mathbf{x}^{(1)} - \mathbf{x}^{(2)})$$

את b מוצאים על ידי בחירת support vector אחד מתוך המשוואה $y^{(i)} (\mathbf{w}^\top \mathbf{x}^{(i)} + b) = 1$.

• נסתכל על הנקודה הראשונה:

$$y^{(1)} (\mathbf{w}^\top \mathbf{x}^{(1)} + b) = 1$$

$$\Leftrightarrow \mathbf{w}^\top \mathbf{x}^{(1)} + b = 1$$

$$\Leftrightarrow b = 1 - \mathbf{w}^\top \mathbf{x}^{(1)} = \frac{\mathbf{w}^\top \mathbf{w}}{\|\mathbf{w}\|_2^2} - \mathbf{w}^\top \mathbf{x}^{(1)} = \mathbf{w}^\top \left(\frac{\mathbf{w}}{\|\mathbf{w}\|_2^2} - \mathbf{x}^{(1)} \right)$$

$$\Leftrightarrow b = \mathbf{w}^\top \left(\frac{1}{2} (\mathbf{x}^{(1)} - \mathbf{x}^{(2)}) - \mathbf{x}^{(1)} \right)$$

$$\Leftrightarrow b = -\frac{1}{2} (\mathbf{x}^{(1)} + \mathbf{x}^{(2)})^\top \mathbf{w}$$

(2) הראו זאת על ידי פתרון הבעיה הפרימאלית.

הבעיה הפרימלית הינה:

$$\begin{aligned} \boldsymbol{w}^*, b^* = \arg \min_{\boldsymbol{w}, b} \quad & \frac{1}{2} \|\boldsymbol{w}\|^2 \\ \text{s.t.} \quad & y^{(i)} \left(\boldsymbol{w}^\top \boldsymbol{x}^{(i)} + b \right) \geq 1 \quad \forall i \end{aligned}$$

- במקרה שבו יש רק נקודה אחת מכל מחלקה, שתי הנקודות בהכרח יהיו **support vectors** ויקיימו $y^{(i)} \left(\boldsymbol{w}^\top \boldsymbol{x}^{(i)} + b \right) = 1$.
- נניח בלי הגבלת הכלליות ש $y^{(1)} = 1$ ו $y^{(2)} = -1$.

שני האילוצים שהנקודות אלו מגדירות הם:

$$\begin{cases} y^{(1)} (\mathbf{w}^\top \mathbf{x}^{(1)} + b) = 1 \\ y^{(2)} (\mathbf{w}^\top \mathbf{x}^{(2)} + b) = 1 \end{cases}$$

$$\Leftrightarrow \begin{cases} (\mathbf{w}^\top \mathbf{x}^{(1)} + b) = 1 \\ -(\mathbf{w}^\top \mathbf{x}^{(2)} + b) = 1 \end{cases}$$

$$\Leftrightarrow \begin{cases} \mathbf{w}^\top \mathbf{x}^{(1)} + b = 1 \\ \mathbf{w}^\top \mathbf{x}^{(2)} + b = -1 \end{cases}$$

$$\Leftrightarrow \begin{cases} b = -\frac{1}{2} (\mathbf{x}^{(1)} + \mathbf{x}^{(2)})^\top \mathbf{w} \\ \mathbf{w}^\top (\mathbf{x}^{(1)} - \mathbf{x}^{(2)}) = 2 \end{cases}$$

נוכל להשתמש באילוץ השני ולכתוב בעזרתו את בעיית האופטימיזציה רק על w :

$$w^* = \arg \min_w \frac{1}{2} \|w\|^2$$

s.t. $w^\top (x^{(1)} - x^{(2)}) = 2$

• בעיה זו מחפשת את ה w בעל האורך המינימאלי כך שהמכפלה הסקלרית שלו עם $(x^{(1)} - x^{(2)})$ היא 2. הוקטור הזה יהיה וקטור בכיוון של $(x^{(1)} - x^{(2)})$ ובאורך של $2/\|x^{(1)} - x^{(2)}\|_2$

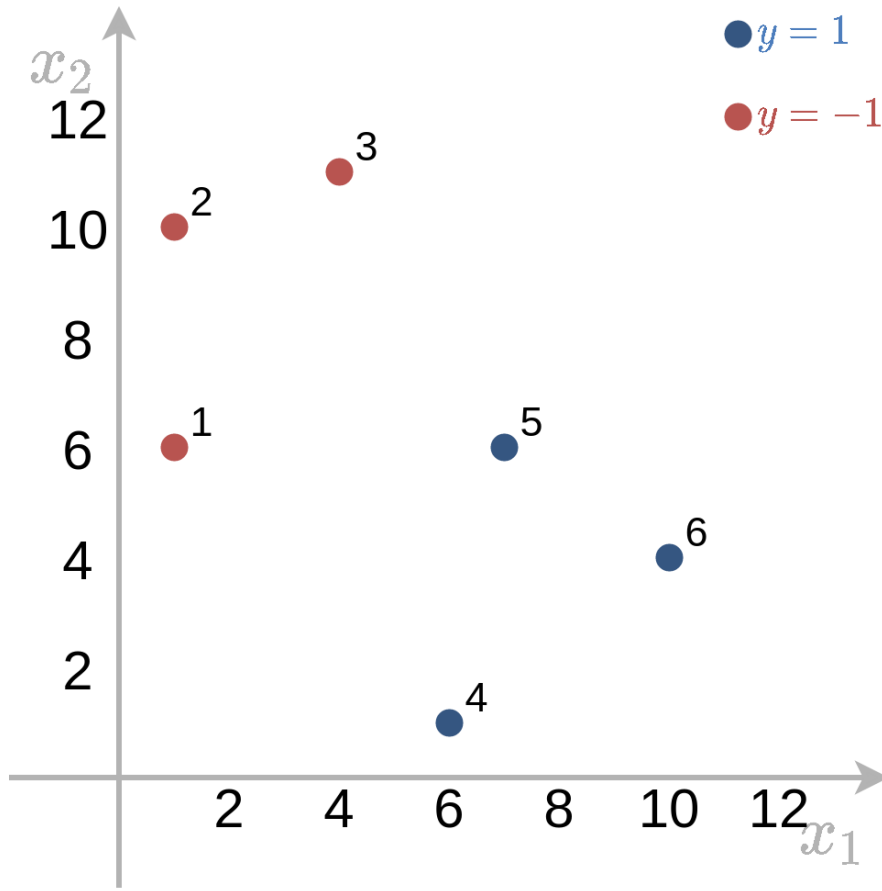
• מכאן:

$$w = \frac{2}{\|x^{(1)} - x^{(2)}\|_2^2} \left(x^{(1)} - x^{(2)} \right)$$

תרגיל 6.2 - Hard SVM

נתון המדגם הבא:

.	1	2	3	4	5	6
y	-1	-1	-1	1	1	1
x ₁	1	1	4	6	7	10
x ₂	6	10	11	1	6	4



1 מצא את מסוג ה hard SVM המתאים למדגם זה. מי הם וקטורי התמיכה?

- בעיית האופטימיזציה הפרימאלית אותה נרצה לפתור הינה:

$$\begin{aligned} w^*, b^* = \arg \min_{w, b} \quad & \frac{1}{2} \|w\|^2 \\ \text{s.t.} \quad & y^{(i)} \left(w^\top x^{(i)} + b \right) \geq 1 \quad \forall i \end{aligned}$$

- באופן כללי פתרון בעיות מסוג זה באופן ידני הוא קשה, אך בעבור מקרים פשוטים נוכל לפתור זאת תוך שימוש בעיקרון הבא:

- במידה ומצאנו את הפיתרון האופטימאלי על תת מדגם מתוך המדגם המלא, ומתקיים שבעבור הפתרון שמצאנו, הנקודות שאינן חלק מתת המדגם נמצאות מחוץ ל margin, אז הפתרון הוא בהכרח הפתרון האופטימאלי בעבור המדגם כולו.

- עקרון זה נכון בגלל העובדה שהוספה של אילוצים לבעיית מינימיזציה יכולה רק להגדיל את ה objective המינימאלי.
- המשמעות היא שניתן לנחש את ה support vectors ולפתור את הבעיה רק בעבור נקודות אלו תוך התעלמות משאר הנקודות.
- לאחר שנפתור את הבעיה על הנקודות שניחשנו יהיה עלינו לבדוק אם שאר נקודות במדגם מחוץ ל margin.
- אם אכן שאר הנקודות מחוץ ל margin אז פתרנו את הבעיה ואם לא אז עלינו לנחש נקודות אחרות.
- בעבור ה support vectors האילוצים של $y^{(i)} (w^T x^{(i)} + b) \geq 1$ הופכים להיות אילוצי שיוון שאיתם הרבה יותר קל לעבוד.

ננסה לנחש מהם ה **support vectors**. נתחיל ב 0 ונגדיל
בהדרגה את כמות ה **support vectors**.

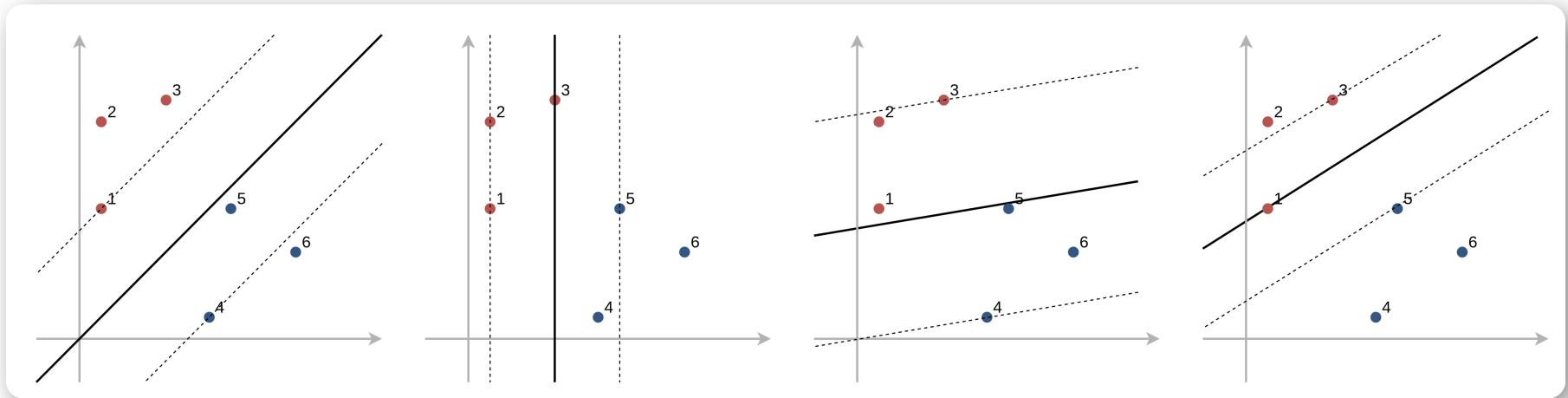
0 or 1 Support Vectors

- באופן כללי בעבור מדגם "נורמלי" אשר מכיל נקודות משני המחלקות שאותן יש לסווג, מקרים אלו לא יוכלו להתקיים.
- באופן תיאורטי במקרה המנוון שבו המדגם מכיל רק תוויות מסוג אחד אז ניתן להגדיר בצורות שונות את כמות ה support vectors, אך מקרים אלו לא מאד רלוונטיים

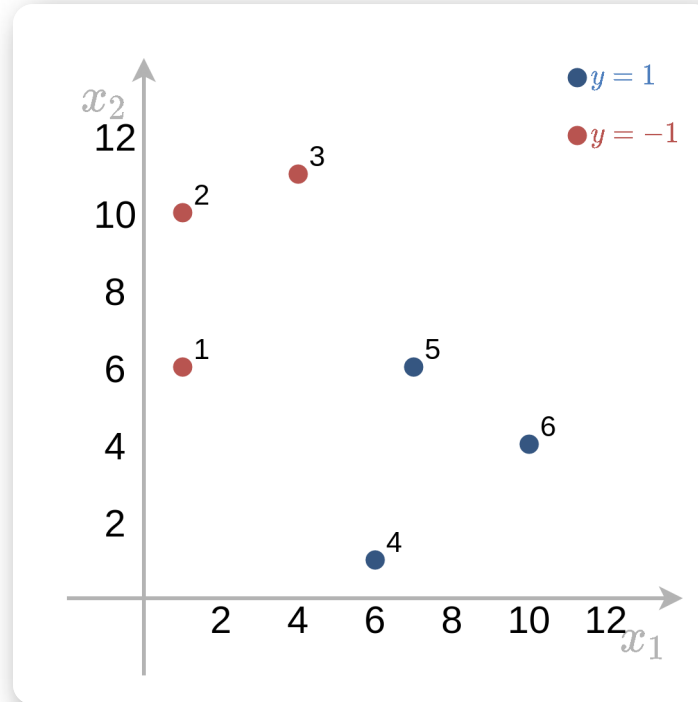
2 Support Vectors

- על פי התרגיל הקודם בעבור שני **support vectors** שמישור ההפרדה יעבור בדיוק במרכז בין 2 הנקודות, יהיה ניצב לוקטור המחבר את שתי הנקודות והשוליים של ה-margin יעברו דרך הנקודות.
- ננסה לחפש 2 נקודות שיכולות להיות **support vectors**.

- נראה שכל בחירה של זוג נקודות יצור איזור של margin שמכיל נקודה אחרת ולכן לא מקיים את האילוצים.
- לכן הפתרון לא יכול להכיל רק שני support vectors.



3 Support Vectors

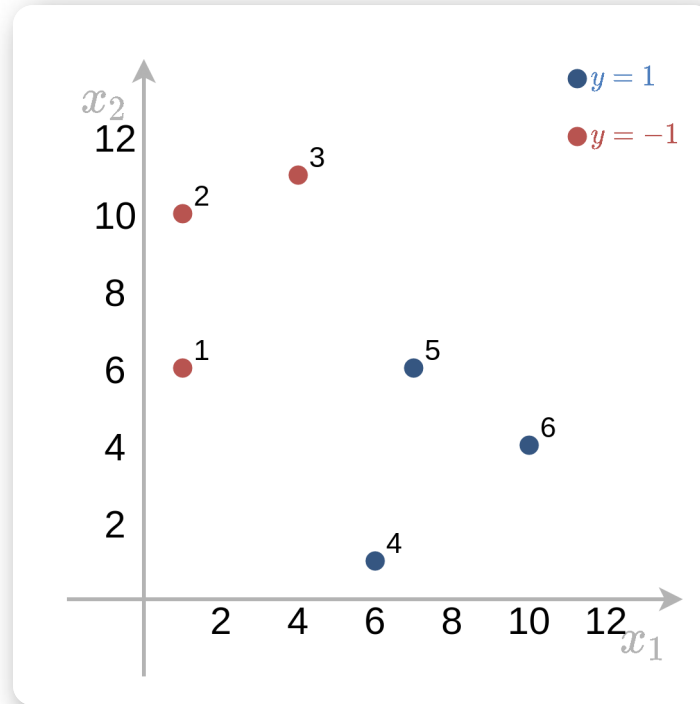


ישנן שתי שלשות שיתכן שיהיו ה support vectors של הפתרון:

$$\{1, 3, 5\} \bullet$$

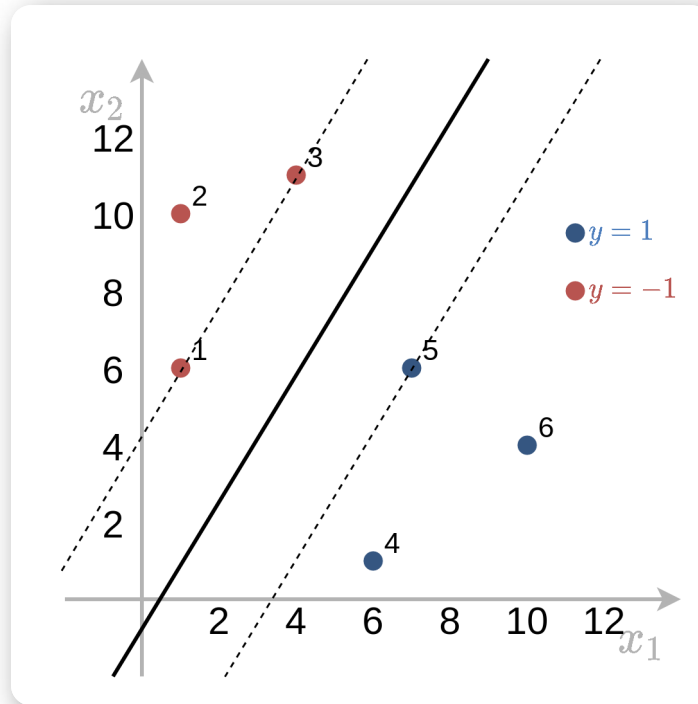
$$\{3, 4, 5\} \bullet$$

- נסתכל ראשית על תת המדגם $\{3, 4, 5\}$.



- נשים לב שכאן הפתרון עבור תת המדגם $\{3, 5\}$ מקיים את האילוץ שגם נקודה 4 מסווגת נכון.
- אולם ראינו שפתרון כזה אינו מקיים את שאר האילוצים!

- השלשה $\{1, 3, 5\}$ מגדירה איזור margin שאינו מכיל נקודות אחרות ולכן מקיים את האילוצים:



- נקודות אלו יהיו שלושת ה support vectors שיגדיר את הפתרון לבעיה. נחשב את הפתרון המתקבל משלושת הנקודות האלה.

מתוך האילוצים נקבל 3 משוואות ב3 נעלמים. בעבור הנקודות $\{1, 3, 5\}$ משוואות אלו יהיו:

$$\begin{cases} -((1, 6)\mathbf{w} + b) = 1 \\ -((4, 11)\mathbf{w} + b) = 1 \\ ((7, 6)\mathbf{w} + b) = 1 \end{cases}$$
$$\Leftrightarrow \begin{cases} w_1 + 6w_2 + b = -1 \\ 4w_1 + 11w_2 + b = -1 \\ 7w_1 + 6w_2 + b = 1 \end{cases}$$

הפתרון של מערכת המשוואות הזו הינה $\mathbf{w} = \frac{1}{15}(5, -3)^\top$ ו $b = -\frac{2}{15}$

2) מבלי לפתור את הבעיה הדואלית. אלו ערכים של $\{\alpha_i\}$ בהכרח יתאפסו?

• בעבור נקודות שאינן **support vectors**, נקבל ש- α_i בהכרח יתאפס

• בבעיה זו הנקודות שאינן **support vectors** הן $\{2, 4, 6\}$ ולכן $\alpha_2 = \alpha_4 = \alpha_6 = 0$.

**(3) מהו הרוחב של ה margin של הפתרון?
ה margin תלוי בגודל של הוקטור w והוא שווה ל:**

$$\frac{2}{\|w\|} = \frac{2 \cdot 15}{\sqrt{5^2 + 3^2}} = 5.145$$

תרגיל 6.3 - גרעין גאوسي

נתון מדגם המכיל 2 נקודות אחת מכל מחלקה:

$$\mathbf{x}^{(1)} = (1, 1)^\top, \quad y^{(1)} = +1$$

$$\mathbf{x}^{(2)} = (-1, -1)^\top, \quad y^{(2)} = -1$$

חשבו את המסווג המתקבל מ **hard SVM** עם גרעין גאوسي
מהצורה $K(\mathbf{x}_1, \mathbf{x}_2) = \exp(-\|\mathbf{x}_1 - \mathbf{x}_2\|^2)$.

פתרון 6.3

הבעיה הדואלית עם הגרעין הגאוסית הינה:

$$\begin{aligned} \{\alpha_i\}^* = \arg \max_{\{\alpha_i\}} & \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y^{(i)} y^{(j)} \alpha_i \alpha_j K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) \\ \text{s.t.} & \alpha_i \geq 0 \quad \forall i \\ & \sum_i \alpha_i y^{(i)} = 0 \end{aligned}$$

בדומה לתרגיל הראשון נקבל ש:

$$\alpha_1 = \alpha_2 = \alpha$$

נחשב את הערכים של $K(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$:

$$K(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) = \exp(0) = 1$$

$$K(\mathbf{x}^{(2)}, \mathbf{x}^{(2)}) = \exp(0) = 1$$

$$K(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = \exp(-(2^2 + 2^2)) = e^{-8}$$

בעיית האופטימיזציה הינה:

$$\begin{aligned}\alpha^* &= \arg \max_{\alpha} 2\alpha - \frac{\alpha^2}{2} \left(K(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) - 2K(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) + K(\mathbf{x}^{(2)}, \mathbf{x}^{(2)}) \right) \\ &\quad \text{s.t. } \alpha \geq 0 \\ &= \arg \max_{\alpha} 2\alpha - \alpha^2(1 - e^{-8}) \\ &\quad \text{s.t. } \alpha \geq 0\end{aligned}$$

נגזור ונשווה ל-0:

$$\begin{aligned}\frac{d}{d\alpha} 2\alpha - \alpha^2(1 - e^{-8}) &= 0 \\ \Leftrightarrow 1 &= \alpha(1 - e^{-8}) \\ \Leftrightarrow \alpha &= \frac{2}{1 - e^{-8}}\end{aligned}$$

הוקטור w נתון על ידי: $w = \sum_i \alpha_i y^{(i)} \Phi(\mathbf{x}^{(i)})$. נכתוב את הביטוי $w^\top \Phi(\mathbf{x})$ בעבור נקודה כל שהיא \mathbf{x} :

$$\begin{aligned} w^\top \Phi(\mathbf{x}) &= \sum_i \alpha_i y^{(i)} \Phi(\mathbf{x}^{(i)})^\top \Phi(\mathbf{x}) \\ &= \sum_i \alpha_i y^{(i)} K(\mathbf{x}^{(i)}, \mathbf{x}) \\ &= \frac{1}{1 - e^{-8}} \left(K(\mathbf{x}^{(1)}, \mathbf{x}) - K(\mathbf{x}^{(2)}, \mathbf{x}) \right) \end{aligned}$$

נחשב את b על ידי שימוש בנקודה הראשונה:

$$\begin{aligned} 1 &= y^{(1)} \left(w^\top \Phi(\mathbf{x}^{(1)}) + b \right) \\ \Leftrightarrow b &= 1 - w^\top \Phi(\mathbf{x}^{(1)}) \\ \Leftrightarrow b &= 1 - \frac{1}{1 - e^{-8}} \left(K(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) - K(\mathbf{x}^{(2)}, \mathbf{x}^{(1)}) \right) \\ \Leftrightarrow b &= 1 - \frac{1}{1 - e^{-8}} (1 - e^{-8}) = 0 \end{aligned}$$

$$\begin{aligned}
h(\mathbf{x}) &= \text{sign}(\mathbf{w}^\top \Phi(\mathbf{x}) + b) \\
&= \text{sign} \left(\frac{1}{1 - e^{-8}} \left(K(\mathbf{x}^{(1)}, \mathbf{x}) - K(\mathbf{x}^{(2)}, \mathbf{x}) \right) \right) \\
&= \text{sign} \left(K(\mathbf{x}^{(1)}, \mathbf{x}) - K(\mathbf{x}^{(2)}, \mathbf{x}) \right) \\
&= \text{sign} \left(\exp \left(-\|\mathbf{x}^{(1)} - \mathbf{x}\|^2 \right) - \exp \left(-\|\mathbf{x}^{(2)} - \mathbf{x}\|^2 \right) \right) \\
&= \text{sign} \left(\|\mathbf{x}^{(2)} - \mathbf{x}\|^2 - \|\mathbf{x}^{(1)} - \mathbf{x}\|^2 \right) \\
&= \text{sign} \left((x_1 + 1)^2 + (x_2 + 1)^2 - (x_1 - 1)^2 - (x_2 - 1)^2 \right) \\
&= \text{sign} (2x_1 + 2x_2) \\
&= \text{sign} (x_1 + x_2)
\end{aligned}$$

חלק מעשי: זיהוי מין הדובר

Code

בחלק זה, ננסה להשתמש ב-SVM כדי לזהות את מינו של הדובר באמצעות קולו. מוטיבציה למערכת כזאת יכולה להיות עוזר וירטואלי שרוצה לפנות לדובר לפי מינו. הרחבה לניסיון זה יכולה להיות זיהוי דובר על סמך קולו וכו'.

הרעיון וה- DATA נלקחו מ- Dataset והערכת ביצועים של קורי בקר, אשר נמצאים באתר הבא. בפרוייקט זה נאספו 3168 דגימות קול מתויגות מהמקורות הבאים:

The Harvard-Haskins Database of Regularly-Timed Speech •

Telecommunications & Signal Processing Laboratory (TSP) Speech Database at McGill University •

VoxForge Speech Corpus •

Festvox CMU_ARCTIC Speech Database at Carnegie Mellon University •

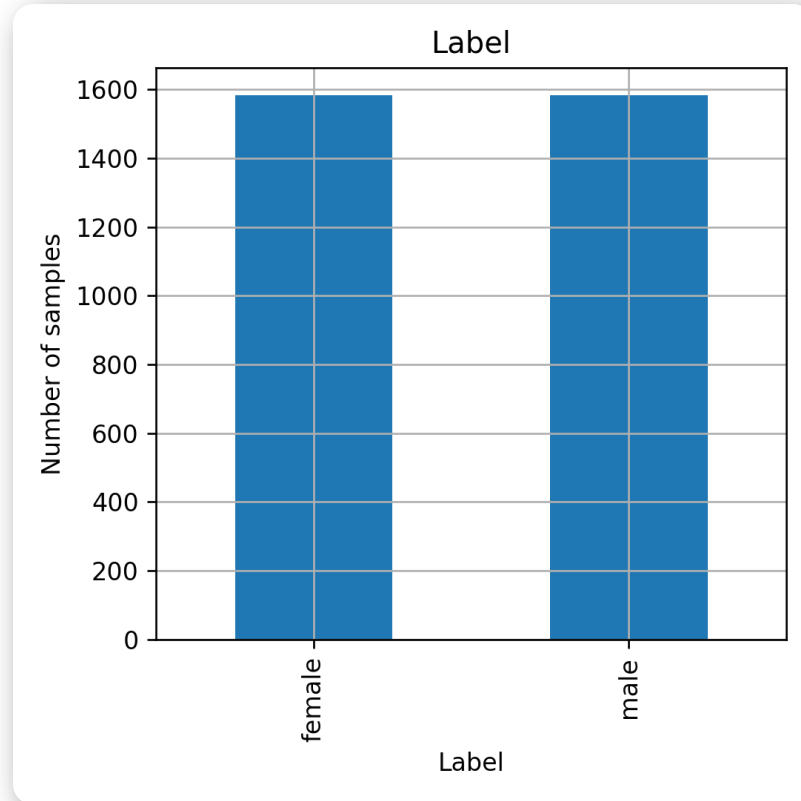
WarbleR כל רצועת קול עברה עיבוד באמצעות כלי בשם
בכדי לייצר 20 Features לכל דגימה:

- **label**: The label of each track: male/female
- **meanfreq**: mean frequency (in kHz)
- **sd**: standard deviation of frequency
- **median**: median frequency (in kHz)
- **Q25**: first quantile (in kHz)
- **Q75**: third quantile (in kHz)
- **IQR**: interquantile range (in kHz)
- **skew**: skewness (see note in specprop description)
- **kurt**: kurtosis (see note in specprop description)

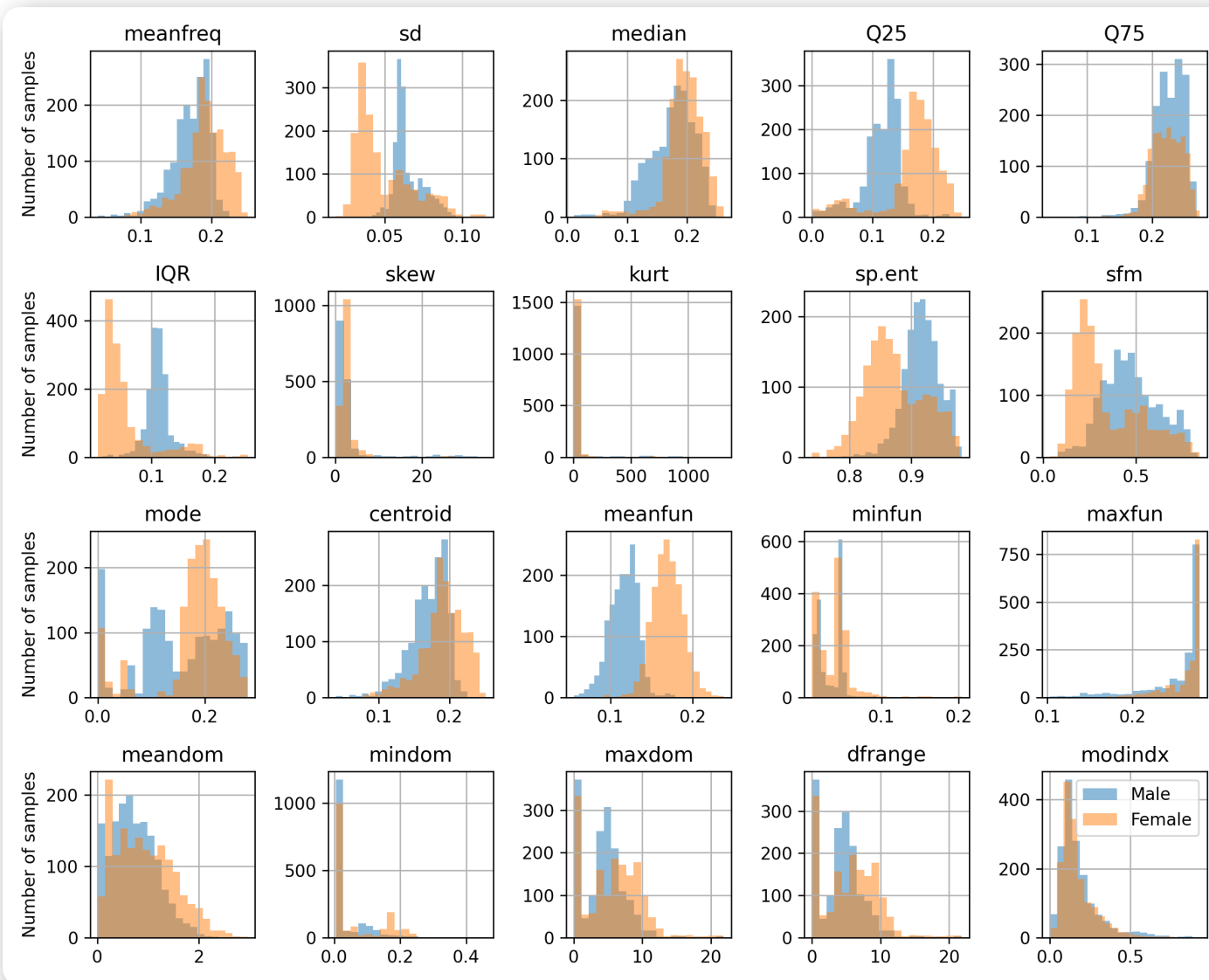
נציג מספר עמודות ושורות מן המדגם:

	label	meanfreq	sd	median	Q25	Q75	IQR	skew	kurt
0	male	0.059781	0.0642413	0.0320269	0.0150715	0.0901934	0.075122	12.8635	274.403
1	male	0.0660087	0.06731	0.0402287	0.0194139	0.0926662	0.0732523	22.4233	634.614
2	male	0.0773155	0.0838294	0.0367185	0.00870106	0.131908	0.123207	30.7572	1024.93
3	male	0.151228	0.0721106	0.158011	0.0965817	0.207955	0.111374	1.23283	4.1773
4	male	0.13512	0.0791461	0.124656	0.0787202	0.206045	0.127325	1.10117	4.33371
5	male	0.132786	0.0795569	0.11909	0.067958	0.209592	0.141634	1.93256	8.3089
6	male	0.150762	0.0744632	0.160106	0.0928989	0.205718	0.112819	1.53064	5.9875
7	male	0.160514	0.0767669	0.144337	0.110532	0.231962	0.12143	1.39716	4.76661
8	male	0.142239	0.0780185	0.138587	0.0882063	0.208587	0.120381	1.09975	4.07028
9	male	0.134329	0.08035	0.121451	0.07558	0.201957	0.126377	1.19037	4.78731

הפילוג של התוויות (גברים נשים) במדגם הינו:



נציג את ההיסטוגרמה של כל אחד מהמאפיינים בעבור כל אחד מתוויות:



Soft SVM

ננסה לפתור את בעיית הסיווג של מין הדובר בעזרת soft SVM. נרצה לשם כך לפתור את בעיית האופטימיזציה הבאה (הבעיה הדואלית):

$$\begin{aligned} \{\alpha_i\}^* &= \arg \max_{\{\alpha_i\}} \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y^{(i)} y^{(j)} \alpha_i \alpha_j \mathbf{x}^{(i)\top} \mathbf{x}^{(j)} \\ \text{s.t.} \quad &0 \leq \alpha_i \leq C \quad \forall i \\ &\sum_i \alpha_i y^{(i)} = 0 \end{aligned}$$

• נמצא את w בעזרת:

$$w = \sum_i \alpha_i y_i x_i$$

- ניתן לחשב את b על ידי בחירה של נקודה שעבורה $0 < \alpha_i$ ולהשתמש במשוואה: $y_i (w^T x_i + b) = 1 - \xi_i$.
- בכדי לפתור את הבעיה נצטרך לבחור פרמטר משקל C אשר קובע את גודל הקנס לנקודות שחורגות מה margin . נתחיל ב $C = 1$ ואחר כך ננסה למצוא את הערך המיטבי.

חלוקת המדגם

כרגיל נחלק את המדגם ל 60%-20%-20% שהם train-validation-test.

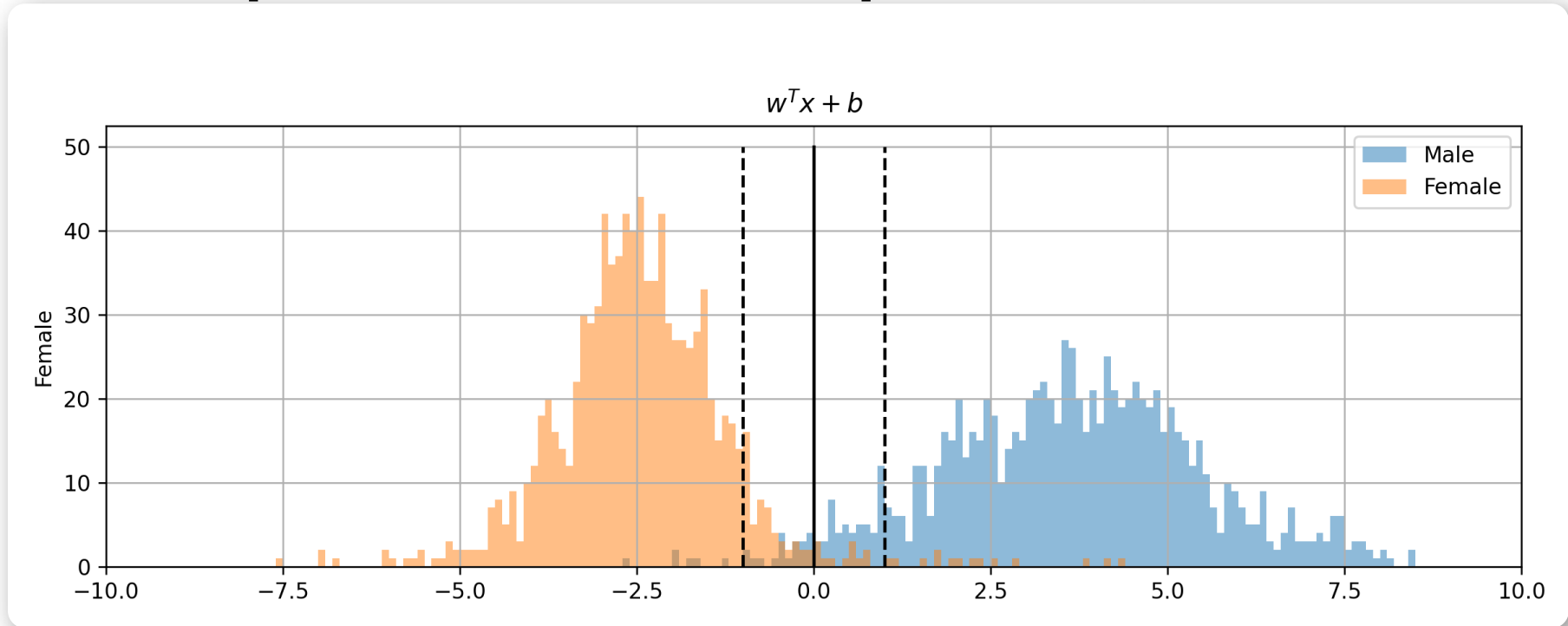
נרמול

חשוב לנרמל את העמודות של המדגם לפני הרצת האלגוריתם משתי סיבות עיקריות:

1. המדגם מכיל מאפיינים ביחידות וסקלות שונות.

2. האלגוריתם מנסה למזער objective אשר מבוסס מרחק, מה שהופך אותו לרגיש לגודל של כל מאפיין. לדוגמא, אם נכפיל מאפיין מסויים בקבוע גדול מ-1 אנו ניתן לו חשיבות יתרה ב objective.

נשתמש באופטימיזציה נומרית על מנת למצוא את הפרמטרים של משטח ההפרדה. לשם המחשה נשרטט את הפילוג של ה signed distance של הנקודות בכל אחת מהמחלקות



ניתן לראות כי המשטח באמת מצליח לחלק את המדגם, עד כדי כמה נקודות שחורגות. ה miscalssification rate שמתקבל על ה test set הינו **.0.028**

מציאת ה C האופטימאלי

נבחן סדרה של ערכי C בתחום 10^{-3} עד 10^3 . כרגיל, בעבור כל ערך נבנה מסוג ונבחן אותו על ה validation set. נקבל את התוצאה הבאה:

