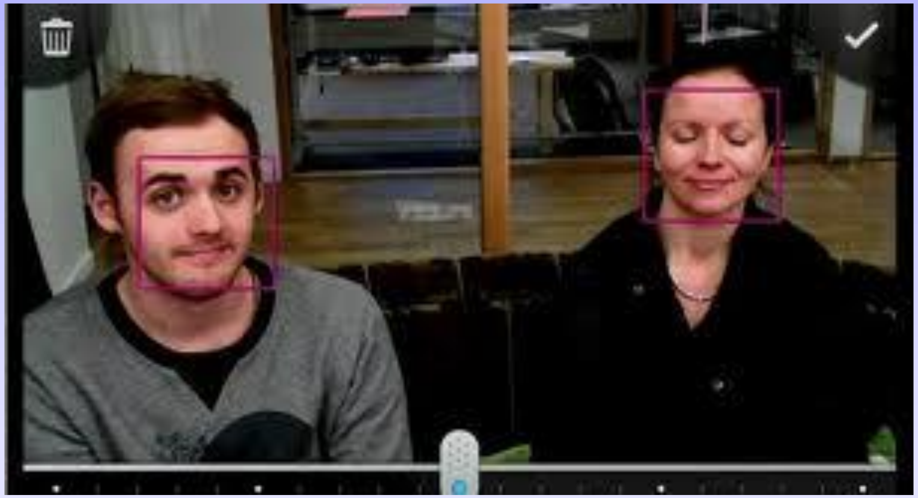
מסווגים ליניאריים

- נתחיל בדוגמא: בקרת גישה אוטומטית לבניין

- עלינו לתכנן מערכת אוטומטית שתאפשר כניסה לבניין מגורים.

דוגמא

- בקרת גישה אוטומטית לבניין
- https://youtu.be/nDhoFBlwjGI



- נעדיף תמונות עם כוון מצלמה-פנים דומה (Camera-face orientation)

דוגמאות מתוייגות

התמונות עבורן ידוע אם האנשים מורשי כניסה או לא, הן תמונות או דוגמאות מתוייגות (labeled images or labeled examples)





"+1" "+1" "+1" "−1" "−1"

תמונה מסומנת כחיובית אם מותר לאדם שהתמונה שלו להיכנס, ומסומנת כשלילית אם אסור לו להיכנס.

מסווגים ליניאריים – המשך

- המטרה: ליצור פונקציה או מסווג (classifier) שימפה את תמונות הפיקסלים (pixel images) לתגיות בינאריות $\{\pm 1\}$
- נתונה הקבוצה של הדוגמאות המתוייגות בלבד – קבוצת האימון.

משימת הסווג

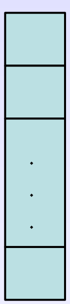
נהפוך את המשימה למעט יותר פורמלית

הנחה: כל תמונה (grayscale) מיוצגת על-ידי וקטור עמודה x ממימד n.

משרשרים את וקטורי העמודות של התמונה לוקטור עמודה אחד ארוך.

לדוגמא: אם התמונה היא 100x100 אזי $n=10,000$.

הנחה: כל התמונות הן בעלות גודל זהה.



משימת הסווג

הנחה: כל תמונה מיוצגת על-ידי וקטור עמודה x ממימד n.

משרשרים את וקטורי העמודות של התמונה לוקטור עמודה אחד ארוך.

לדוגמא: אם התמונה היא 100x100 אזי n=10,000.

$$\begin{pmatrix} x_{1,1} & x_{1,2} & & x_{1,100} \\ x_{2,1} & x_{2,2} & & x_{2,100} \\ & & & \\ x_{100,1} & & & x_{100,100} \end{pmatrix} \Rightarrow x = \begin{pmatrix} x_1 \\ x_2 \\ \\ x_{10,000} \end{pmatrix}$$

משימת הסווג

הנחה: כל התמונות הן בעלות גודל זהה.

המסווג שלנו הוא פונקציה בינארית:

$$f : R^n \to \{-1, 1\}$$

הנקבעת על בסיס קבוצת האימון בלבד.

אנו מניחים שהמסווג של משימה זו אינו יודע דבר על תמונות (או על תמונות פנים לצורך העניין) מלבד נתוני קבוצת האימון עם התיוגים שלהם.

וקטורי קבוצת האימון יכולים באותה מידה להיות מדידות של משקל, גובה, וכו', או איפיון של אתרי אינטרנט, או מדדים כמו גובה המשכורת, חוב, סכום ההפקדה, וכו'.

לימוד מדוגמאות

הקלט למסווג : קבוצה של m וקטורי אימון :

$$\underline{x}^{(1)}, \underline{x}^{(2)}, \ldots, \underline{x}^{(m)}$$

עם התגיות המתאימות :

$$y^{(1)}, y^{(2)}, \ldots, y^{(m)}$$

$$\underline{x}^{(i)} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \\ x_n \end{pmatrix}$$

$"+1"$

$"+1"$

$"+1"$

$"-1"$

$"-1"$

איזה סוג של פתרון מספיק?

- נניח m=50 דוגמאות מתוייגות, עם פיקסלים ברמות אפור מ-0 עד 255. $\{x^{(1)},y^{(1)}\},\{x^{(2)},y^{(2)}\},\ldots,\{x^{(m)},y^{(m)}\}$
- אפשר למצוא פיקסל בודד (רכיב בודד בוקטור התכונות), כך שערכו שונה עבור כל אחת מ- m התמונות.
- נסמן: $x_j^{(t)}$ - הפיקסל ה- j של התמונה ה- t
  x'_j - הפיקסל ה- j של התמונה x'
- נציע מסווג בינארי:

$$f_i(x') = \begin{cases} y^{(t)} & if \quad x^{(t)}{}_j = x'_j \\ -1 & otherwise \end{cases}$$

איזה סוג של פתרון מספיק?

- מסווג זה יפתור את הבעייה באופן מושלם.

$$f_i(x') = \begin{cases} y^{(t)} & if \quad x^{(t)}{}_j = x'_j \\ -1 & otherwise \end{cases}$$

- האם פתרנו את הבעייה?

האם הכלל נכון עבור תמונות שלא נמצאות בקבוצת האימון?






רוצים למצוא מסווגים שמבצעים הכללה היטב, כלומר מסווגים שתיפקודם על קבוצת האימון מייצג עד כמה הם יעבדו היטב עבור דוגמאות חדשות

מסווגים ליניאריים דרך הראשית

- עתה נקבע את מחלקת הפונקציות:
- נתחשב רק במסווגים ליניאריים.

$$f(x,\theta) = sign(\underline{\theta}^T \underline{x})$$

- הוקטור θ: וקטור עמודה עם פרמטרים ממשיים

$$\theta = (\theta_1, \theta_2, \ \dots \ \theta_n)^T$$

$$\theta \in R^n$$

פרמטרים שונים יוצרים פונקציות שונות במחלקה זו, כלומר פונקציות שהיציאה שלהן $\{-1,1\}$ עשויה להיות שונה עבור ערכי קלט x.

תיאור גאומטרי

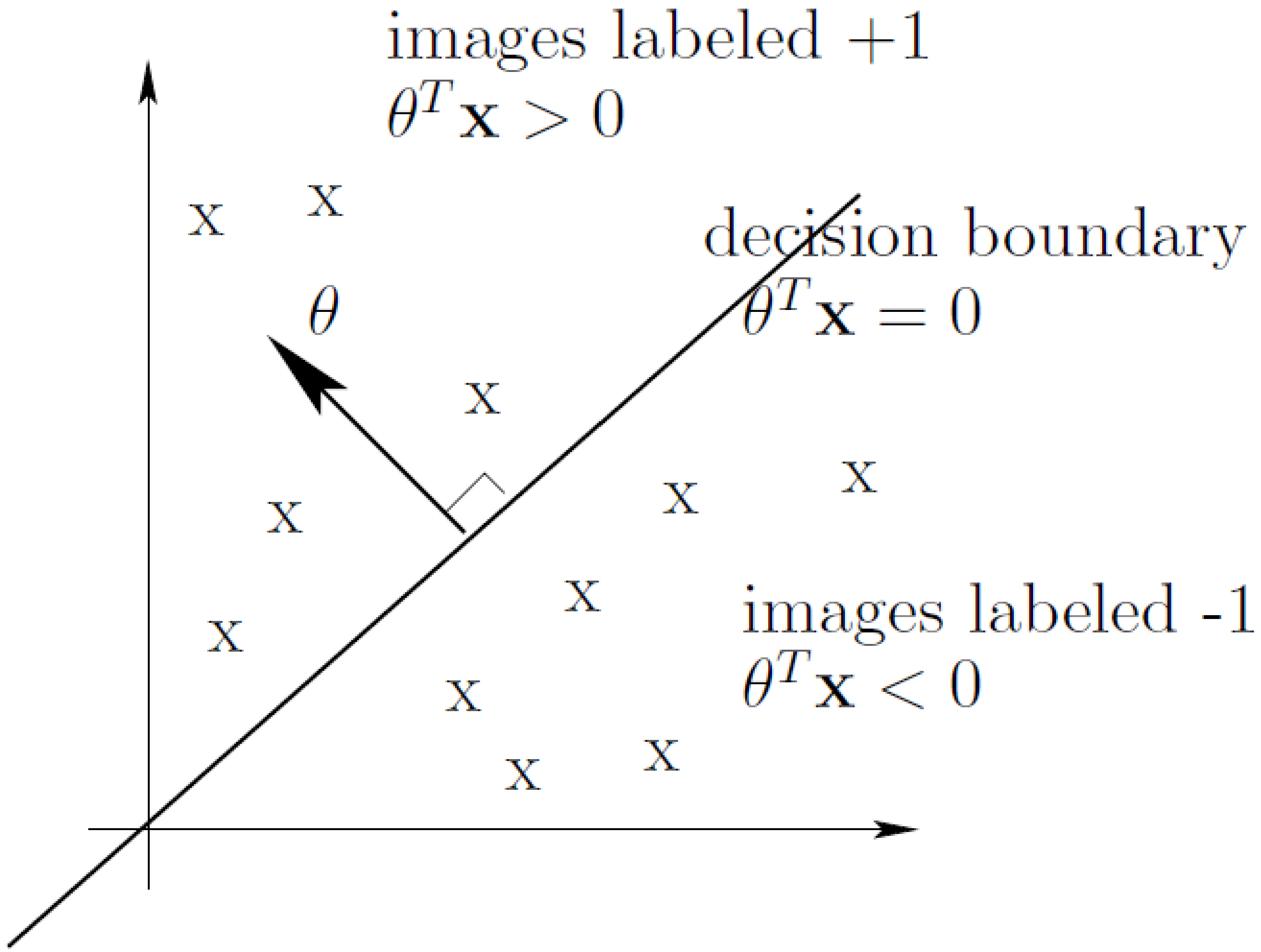
- המסווג משנה את החיזוי רק כאשר הארגומנט של פונקציית הסימן משתנה מחיובי לשלילי או ההפך.
- באופן גאומטרי, במרחב וקטורי התכונות (או התמונות) מעבר זה מתאים לחציית "גבול החלטה" או "משטח החלטה" (Decision Boundary) בו הארגומנט הוא בדיוק אפס:

$$\theta^T \underline{x} = 0$$

$$(\theta_1 x_1 + \theta_2 x_2 + \quad \ldots + \theta_n x_n = 0)$$

- משוואה זו מגדירה מישור במרחב n מימדי, העובר דרך הראשית (כי וקטור האפס פותר את המשוואה).

תיאור גאומטרי



אלגוריתם למידה - הפרספטרון

- לאחר בחירת מחלקת הפונקציות – רוצים לבחור אחת ספיציפית במחלקה שתבצע את פעולת הסווג היטב על קבוצת האימון
- בעייה זו מכונה בעיית השערוך (Estimation problem).
- מספר השגיאות המינימלי על קבוצת האימון, כלומר θ הממזער את שגיאת האימון:

$$E(\theta) = \frac{1}{m}\sum_{i=1}^{m} \left(1-\delta\left(y^{(i)}, f\left(\theta^T x^{(i)}\right)\right)\right)$$

- כאשר:

$$\delta(y, y') = \begin{cases} 1 & if \quad y = y' \\ 0 & otherwise \end{cases}$$

אלגוריתם למידה - הפרספטרון

- פונקציית שגיאת האימון מונה את המספר הממוצע של שגיאות – כלומר המספר הממוצע של תמונות או וקטורי אימון בהן הפונקציה של המסווג חוזה תיוג שונה מהתיוג הנכון (הידוע של התמונה הנתונה.

$$E(\theta) = \frac{1}{m} \sum_{i=1}^{m} (1 - \delta(y^{(i)}, f(\theta^T x^{(i)})))$$

- במונחים של פונקציית עלות $\mathcal{L}$: $\mathcal{L}=1$ כאשר תיוג האימון שונה מתיוג פונקציית המסווג.
- פונקצייה כללית $\mathcal{L}$ יכולה להביא בחשבון גם ערכי עלות שונים

(אדם לא מורשה הנכנס לבניין ומבצע הרס רב או פועל הניקיון שהמערכת מסווגת אותו באופן מוטעה)

- לצורך הפשטות נשתמש בפונקציית עלות אפס-אחד: 1- עבור שגיאות ו- 0 אחרת.

אלגוריתם למידה - הפרספטרון

- מהו אלגוריתם למידה סביר לקביעת הפרמטרים θ?
- אולי אפשר לכוון ולתקן את הפרמטרים באופן אינקרמנטלי לפי השגיאות שהמסווג מבצע. אלגוריתם כזה יפחית את שגיאת האימון.
- מתחשבים בכל תמונה או וקטור מקבוצת האימון בזה אחר זה באופן מחזורי דרך כל נתוני האימון ומעדכנים את הפרמטרים לפי חוק העדכון:

$$if \quad y^{(i)} \neq f(x^{(i)},\theta) \quad then \quad \theta' \leftarrow \theta + x^{(i)} \cdot y^{(i)}$$

- במלים אחרות הפרמטרים משתנים רק אם מבצעים שגיאה

אלגוריתם למידה - הפרספטרון

- עדכונים אלה נוטים לתקן שגיאות.
- כדי לראות את זה, נציין כי כאשר מתבצעת שגיאה:

$$sign(\theta^T x) \neq y^{(i)}$$

- ואז:

$$y^{(i)}\theta^T x < 0$$

- המכפלה חיובית עבור תמונות או וקטורים המסווגים נכון.

$$y^{(i)}\theta^T x > 0$$

אלגוריתם למידה - הפרספטרון

- נניח שמבצעים שגיאה על $x^{(t)}$
- אזי וקטור הפרמטרים מעודכן באופן הבא:

$$\theta' = \theta + x^{(t)} y^{(t)}$$

- נתאר את הסווג של התמונה לאחר העדכון:

$$y^{(t)} \theta'^T x^{(t)} = y^{(t)} (\theta + x^{(t)} y^{(t)})^T x^{(t)} = y^{(t)} \theta^T x^{(t)} + y^{(t)2} \left(x^{(t)} \right)^T x^{(t)}$$

$$= y^{(t)} \theta^T x^{(t)} + \| x^{(t)} \|^2$$

אם נסווג את התמונה הזו שוב ושוב, אז בהכרח נשנה את הפרמטרים כך שהתמונה תסווג באופן נכון, כלומר שהערך של המכפלה $y_t \theta^T x_t$ יהיה חיובי.

אנליזה של אלגוריתם הפרספטרון

אלגוריתם הפרספטרון מסיים לבצע את עדכון הפרמטרים רק כאשר כל תמונות האימון מסווגות נכון (אין שגיאות – אין עידכונים).

אם אפשר לסווג את תמונות האימון באופן נכון באמצעות מסווג ליניארי, האם הפרספטרון ימצא מסווג כזה?

אנליזה של אלגוריתם הפרספטרון

כן, ובמספר סופי של צעדים.