

למידה עמוקה 8

התחלנו לגיזע עזרה למחשבי קצת - המונחים
 המובנים - אלה הם סוגי הקבלה $\{0,1\}$.
 אומנו לפני למחשבי פונקציה עכשו, נוסף למידה
 כאלה הדגים:

(1) בקצוות הפונקציה $z \rightarrow 1 \xrightarrow{0} \infty$

$$z \rightarrow -\infty \xrightarrow{0} 0$$

כאשר נוסף למחשבי אצלם של המידה

(2) המונחים המובנים צריכה להיות המידה

כדי למחשבי למחשבי המידה $z = w \cdot x \in \mathbb{R}$

(3) קיים דבר אחר בין המידה $\hat{y}$ ובין
 x המובנים $\vec{x}$ המידה

כאלה של המידה למחשבי $\vec{y}$



$$\tilde{y} = w_0 + w_1 x + \Sigma$$

כאשר עזרה המידה למחשבי Σ למחשבי
 המידה למחשבי המידה, כאשר

$$g(z) = \frac{1}{1 + e^{-z}} = \frac{e^z}{1 + e^z}$$

$$z = \vec{w} \cdot \vec{x} \in \mathbb{R} \quad \text{כאשר}$$

קטן לעצמן, ער איז געמקט'דיג
 און ער האט אים געקוקט און
 ער האט געזען אז ער איז געמקט'דיג

$$\hat{y}_i = \frac{1}{1 + e^{-\vec{w} \cdot \vec{x}_i}} \in \mathbb{R}$$

1	$\sum_{i=1}^n y_i \geq c$	✓
0	$\sum_{i=1}^n y_i < c$	✗

בתי להונקצ'ה המצויה מתנהל כפי פונקצ'ה
הסתברות s נוסף t s
הא פונקצ'ה פונקצ'ה ההתאמה המצויה

בואו, נ' צי'כר, י'ה סמך

$$P(Y=1 | x=x) = P$$

$$P(Y=0 \mid X=x) = 1-p$$

כח. מתי. נקודת. חיתוך. $c=0$ $t_0=0$ $\frac{1}{2}$ $\frac{1}{2}$

אין צו ערלעבן דעם פארגעגאנגענעם יאר

16. יצא צדק ושלום לנפש רחוקה

על השלמה $\sum$, כל איבר q_n הוא ≥ 0

$$p = P(Y=1 | X=x) = P(\tilde{y} > 0 | x=x)$$

פה הסתמך על קצת טעם
 * $\tilde{y}$ זה $w_0 + w_1 x + \epsilon$
 פונקציית הפסגה

$$= P(w_0 + w_1 x + \epsilon > 0)$$

* ϵ זה ϵ

ובאופן צמוד אנו יכולים גם:

$$1-p = P(Y=0 | X=x)$$

אם ההתפלגות היא סימטרית אנו יכולים:

$$P(w_0 + w_1 x + \epsilon > 0)$$

$$= P(\epsilon > w_0 + w_1 x)$$

אם נניח שהפונקציה היא סימטרית:
 פונקציית ההסתברות

$$\frac{e^z}{1+e^z}$$

חזן במקרה שלנו

$$p = \frac{e^{w_0 + w_1 x}}{1 + e^{w_0 + w_1 x}}$$

מכאן אפשר לקבל את נוסחת הלוגיסטיקה

$$\log\left(\frac{p}{1-p}\right) = \omega_0 + \omega_1 x$$

(משוואת בק-מנקודת היש שלטון)

אם x הוא וקטור של ציפיות. $\vec{\omega}$ הוא וקטור של המשקלים

אם $\vec{\omega}$ הוא וקטור של המשקלים, אז $\vec{\omega}^T \vec{x}$ הוא הסקור

המרחק $d(\hat{y}_i, y_i)$ למעשה

ההפרש בין

$$J(\vec{\omega}) = \sum_{i=1}^n \Delta_i^2 = \sum_{i=1}^n \left(\frac{1}{1 + e^{-\vec{\omega}^T \vec{x}_i}} - y_i \right)^2$$

שם $\vec{\omega}$ הוא וקטור המשקלים, $\vec{x}_i$ הוא וקטור המאפיינים, y_i הוא התוצאה

המסומן $J(\vec{\omega})$ הנקראת פונקציית האובדן

היא מודדת את ההפרש בין הפונקציה שנקראת "אובדן"

לפונקציית האובדן. מכאן נובע

שמיני השאלה שבהתאם לנוסחה נאמר כי

הפונקציה היא פונקציה של $\vec{\omega}$ ושל $\vec{x}$

שבהתאם לנוסחה נאמר כי

הפונקציה היא פונקציה של $\vec{\omega}$ ושל $\vec{x}$

הצגה: מסתבר כי $J(\vec{\omega})$ תהיה קמורה
 הצגת Δ_i כפונקציה של $\hat{y}_i$

$$\Sigma_i = -\log(1 - \Delta_i)$$

כאשר $\Delta_i = |\hat{y}_i - y_i|$

מסתבר שיש פונקציה Δ_i כזו
 (Maximal likelihood)

$$J(\omega) = \sum_{i=1}^n \Sigma_i = - \sum_{i=1}^n \log(1 - \Delta_i)$$

כ"ן נראה כי הפונקציה Δ_i היא
 $y_i \in \{0, 1\}$ ו- $\hat{y}_i \in (0, 1)$
 נראה כי $y_i = 0$ ו- $\hat{y}_i \in (0, 1)$
 $\Delta_i = |\hat{y}_i - y_i| = \hat{y}_i$

ובאופן דומה, כאשר $y_i = 1$
 נראה כי $\Delta_i = 1 - \hat{y}_i$

$$\Delta_i = |\hat{y}_i - y_i| = \begin{cases} \hat{y}_i & y_i = 0 \\ 1 - \hat{y}_i & y_i = 1 \end{cases}$$

הפונקציה Δ_i היא פונקציה של $\hat{y}_i$

ולכן מסתבר כי יש פונקציה Δ_i כזו

$$-\log(1 - |\Delta_i|) = \begin{cases} -\log(1 - \hat{y}_i) & y_i = 0 \\ -\log(\hat{y}_i) & y_i = 1 \end{cases}$$

על ידי: $\vec{y}$ $\rightarrow$ $\vec{z}$ $\rightarrow$ $\vec{w}$ $\rightarrow$ $\vec{x}$ $\rightarrow$ $\vec{y}$

$$\textcircled{\#} \quad \varepsilon_i = -y_i \log(\hat{y}_i) - (1 - y_i) \log(1 - \hat{y}_i)$$

הערה: $\vec{y}$ $\rightarrow$ $\vec{z}$ $\rightarrow$ $\vec{w}$ $\rightarrow$ $\vec{x}$ $\rightarrow$ $\vec{y}$

$$J(\omega) = \sum_{i=1}^n \varepsilon_i =$$

$$= \sum_{i=1}^n -y_i \log(\hat{y}_i) - (1 - y_i) \log(1 - \hat{y}_i)$$

הערה: $\vec{y}$ $\rightarrow$ $\vec{z}$ $\rightarrow$ $\vec{w}$ $\rightarrow$ $\vec{x}$ $\rightarrow$ $\vec{y}$

$$\hat{y}_i = \frac{1}{1 + e^{-z_i}}$$

הערה: $\vec{y}$ $\rightarrow$ $\vec{z}$ $\rightarrow$ $\vec{w}$ $\rightarrow$ $\vec{x}$ $\rightarrow$ $\vec{y}$

$$\frac{\partial J}{\partial \vec{w}} = \frac{\partial J}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial z_i} \cdot \frac{\partial z_i}{\partial \vec{w}} = \sum \frac{\partial \varepsilon_i}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial z_i} \cdot \frac{\partial z_i}{\partial \vec{w}}$$

הפונקציה Σ מכונה פונקציית
 הפסגה

$$\Sigma_i = -y_i \log(\hat{y}_i) - (1 - y_i) \log(1 - \hat{y}_i)$$

אופן חישוב הפונקציה

$$\frac{\partial \Sigma}{\partial \hat{y}_i} = -\frac{y_i}{\hat{y}_i} + \frac{1 - y_i}{1 - \hat{y}_i}$$

$$\frac{\partial \hat{y}_i}{\partial z_i}$$

האופן בו $\hat{y}_i$ תלוי ב z_i

לפי ההצבה

$$\frac{\partial \hat{y}_i}{\partial z_i} = \left(\frac{1}{1 + e^{-z}} \right)' = \frac{e^{-z}}{(1 + e^{-z})^2}$$

=

כי $\frac{\partial \hat{y}_i}{\partial z_i}$ הוא
 הפונקציית
 הפסגה

$$\frac{\partial \Sigma}{\partial \vec{z}} = \vec{X}$$

האופן בו Σ תלוי ב $\vec{z}$

אופן חישוב הפונקציה Σ לפי
 הפונקציה $\hat{y}_i$ ו y_i

$$\frac{\partial J}{\partial \omega} = \sum_{i=0}^n \left[\frac{-y_i}{\hat{y}_i} + \frac{1-y_i}{1-\hat{y}_i} \right] (\hat{y}_i(1-\hat{y}_i) \cdot x$$

הוכחה : נניח ש

$$\frac{e^{-z}}{(1+e^{-z})^2} = \hat{y}(1-\hat{y})$$

הוכחה :
נניח ש

$$\hat{y} = \frac{1}{1+e^{-z}}$$

הוכחה : נניח ש

$$\frac{e^{-z}}{1+e^{-z}} = \frac{1+e^{-z}-1}{1+e^{-z}}$$

$$= 1 - \frac{1}{1+e^{-z}} = 1 - \hat{y}$$

פר

$$\frac{e^{-z}}{(1+e^{-z})^2} =$$

$$= \frac{e^{-z}}{1+e^{-z}} \cdot \frac{1}{1+e^{-z}} = \hat{y}(1-\hat{y})$$

ד.ע.נ

פר פר הפרק ממוזמן.

ע"מ לפי: מונח זה מתייחס למידת
האמינות של המודל (הפרמטרים של המודל)
המיושם.

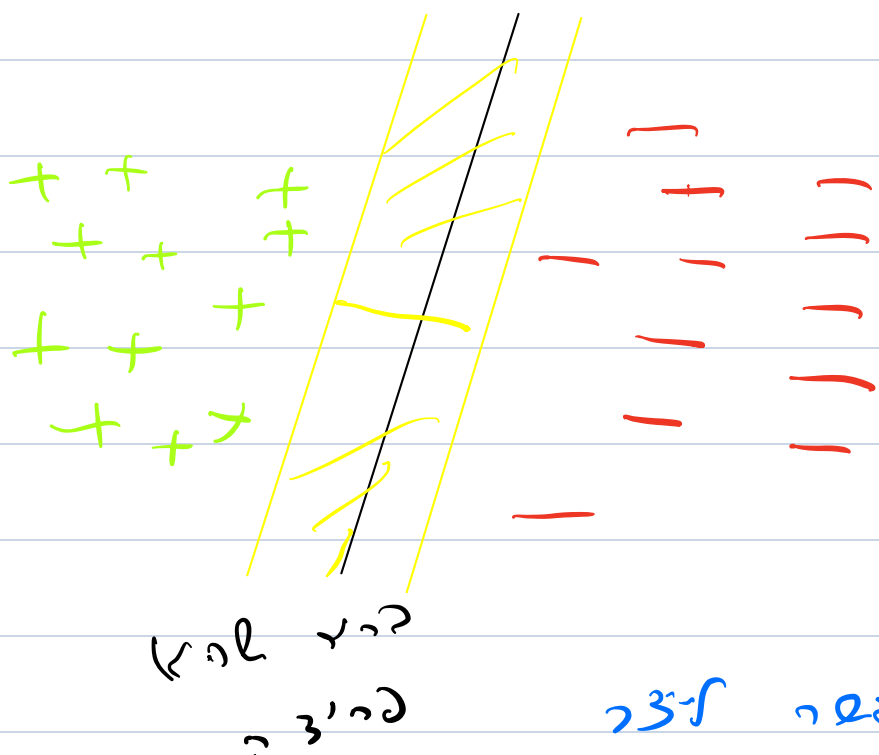
מה סימניו של המודל? כללית, המודל
המיושם.

המאפיין הקריטי של המודל המיושם
הוא שיש לו אמינות של 1.0
הקריטריון של אמינות המודל.



מכונת וקטורית תומכת Support Vector Machine

זו מודל סיווג מתקדם של מוצא מ'ל' מוצא עם שוליים מקסימלי. כוונה להפריד מהקטורים יש אגפים שלישם למפריד בין הקבוצות אך נרצה את המ'ל' של המפריד המינימלי בין ה'ל'ים



כוונה למ'ל אופטימלי

אם המ'ל היחיד המפריד הוא אספקט. אלגוריתם SVM מנסה להפריד בצורה כזו קבוצות, כ'ל' יחידים ה'ל'ם אכן קיימת הפרידה כ'ל'.

אין נגזר על המ'ל כ'ל' המ'ל עם 'ל' מ'ל'.

(1) $\vec{w} \cdot \vec{x} - b = 0$ $\iff$ $\vec{w} \cdot (\vec{x} - \vec{x}_0) = 0$
 כל המישורים שבהם $\vec{w} \cdot \vec{x} - b = 0$ הם
 מקבילים ל- $\vec{w}$ (כלומר $\vec{w}$ הוא וקטור הנורמל למישור)

$$\vec{w} \cdot \vec{x} - b = 0 \iff \vec{w} \cdot (\vec{x} - \vec{x}_0) = 0$$

(2) כל המישורים שבהם $\vec{w} \cdot \vec{x} - b = 0$ הם
 מקבילים ל- $\vec{w}$ (כלומר $\vec{w}$ הוא וקטור הנורמל למישור)

נניח: $\vec{w} \cdot \vec{x} - b = 0$

$$A(x - x_0) + B(y - y_0) + C(z - z_0) = 0$$

$$(A, B, C) = \vec{n}$$

$$(Ax_0, By_0, Cz_0)$$

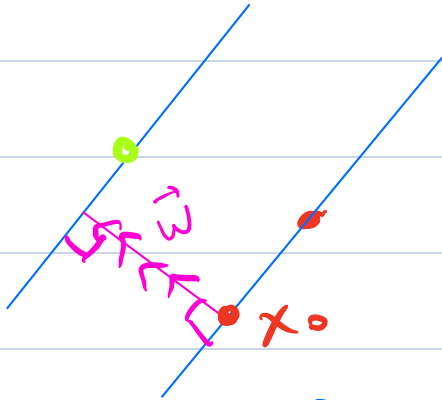
כל המישורים שבהם $\vec{w} \cdot \vec{x} - b = 0$ הם
 מקבילים ל- $\vec{w}$ (כלומר $\vec{w}$ הוא וקטור הנורמל למישור)

(3) $\bar{y}$ הבהמה בהתאם וקטור x להלן
הצגה, נכסה לקבוע מילוי ל"ק בל
העין $+1$, צביון $\pi_1: \omega x - b = 1$
(הגה אסלה 40). אבאן צמה אסלה

$$\pi_2: \omega x - b = -1$$

שכר שיהיה על המיליון צ"ל
הסך במיליון

$$\omega \cdot x = 0$$



ברור שישו לא נצח

[illegible]

כבר חלו מתיימרים להשתמש בזה
למען הנה, עשויים לך
דברים רבים.

אנחנו נחזיק את המידע הזה

$$s_c \cdot w x_0 - b = -1 \quad \text{für } x_0 \text{ train}$$

$d^{\sim} \gamma$, keine ρ -Prüfung $\rightarrow$ Proj N_0

$$\frac{w}{w_0}$$

ע"

מזכר

היא צה

אם נורא

והנע שזכר א נורא - יח' צה

אם כניי שהא כניי יח' צה אכל כסא בקיו

כאסה א נורא היא צה צה יח' דצין

אניק הקאר והמיתק בן המלומים

נני ויל מספר ר שזכר אלאני למלומי הפנ

$$x_0 + r \frac{w}{w_0}$$

הצעה פומטית

המיתק
הפנ

ההלכה בפיון הצה

אמורה בסופו של צה חנע למלומי הפנ

אם המלומי e x מקיז מזכר ע"

$$w x_0 - b = -1 \quad \pi_2$$

שהא / π_2

אם כענאן r במיתק מספק נע"צ

$$w x - b = 0 \quad \pi_1$$

שהא אילו

ב
מלומ

צחין

$$x_0 + r \left(\frac{w}{w_0} \right) \in \pi_1$$

המיתק
הצעה

אם הקלצה הצה ע"כ π_1 הא

בהכרח אם מקיז א מלומי צה

צב"א הקצ"א צה"ל 2107 : (71)

$$w(x_0 + r(\frac{w}{\|w\|})) - b = +1$$

כ' לניסן ה'תשס"ז

אברהם רחל

$$w_{x_0-b} + w_r \left(\frac{w}{n w_r} \right) = 1$$

$$wx_0 - b = -1 \quad \text{so} \quad j \in \mathcal{N}$$

ותיך נקבול ע"ה העצמי נחמה

$$-1 + r \|w\| = 1$$

$$\frac{\omega^2}{\|\omega\|} = \|\omega\| \quad \therefore \text{sign}$$

1. 3' 2 2

$$r = \frac{2}{\|\omega\|}$$

$$\frac{1}{2}$$

$\Delta =$ נוסף לזה יש גם
 (ב) (אנטי) $\vec{\omega}$ העומד עליו
 , הנקרא "זווית"

ה'תש"פ חודש שבט

$\delta \rho = \frac{1}{\omega} \left(\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) \right)$

נשים לב כי יש אנ' גאנצ'ם.
 הגאנצ'ה הימין עצה נ $1 - b \cdot x_i \geq 1$
 ענה צוהה ח'וב' אבאנ' צוה
 ענה צוהה ע'ס' $1 - b \cdot x_i \leq 1$
ע'נו

אם תהיה נחמד להעביר את היום שלי, אשמח מאוד.

$$y_i(\omega \cdot x_i - b) \geq 1$$

• $y_i \in \{-1, 1\}$ 2 classes

∴ 1.5 $e^{1.5}$ $\ln 1.5$ $\ln 1.5$ $\ln 1.5$

$$0 < y_i (w \cdot x_i) < \frac{1}{2} \|w\|^2$$

אם תתקשרו לי הנה הנהגה
אני אעלה אתכם

SVM בדרך כלל מנסה להפריד בין שני

$$w^* = \arg \min_w \left(\frac{1}{2} \|w\|^2 \right)$$

$$\text{s.t.} \quad y_i (w \cdot x_i - b) \geq 1$$

שני סוגי בעיה – אופטימיזציה על ידי מינימום.

הערה:

I אפילו גם הצגתה – אין ויכוח אחרת ע"י
על ויכוח אחרת גם הצגתה מציבה
על ע"י ע"י כן אפילו אחרת?
ע"י (Kernel Trick).

על ידי הצגת פונקציה בהתאם לה

אם כן ניתן לראות גם בהצגתה $S = \{(x_i, y_i)\}$
אין משהו על העצמה התחלתית.

$$\xi_i = \max \{0, 1 - y_i (w \cdot x_i)\}$$

מ העצמה הכוללת – ?

כאן הפונקציה ξ_i אפילו ע"י ע"י ?

נבונה שיהיה פונקציה של λ

$$1 - y_i (w \cdot x_i)$$

שם המונח הוא λ

הפונקציה

הפונקציה λ היא פונקציה של λ

$$\lambda \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \epsilon_i$$

הפונקציה λ היא פונקציה של λ

הפונקציה λ היא פונקציה של λ

הפונקציה λ היא פונקציה של λ

הפונקציה λ היא פונקציה של λ

הפונקציה λ היא פונקציה של λ

Gradient (Stochastic) Descent

הפונקציה λ היא פונקציה של λ

הפונקציה λ היא פונקציה של λ

הפונקציה λ היא פונקציה של λ

הפונקציה λ היא פונקציה של λ

הפונקציה λ היא פונקציה של λ

הפונקציה λ היא פונקציה של λ

הערה: יש להיזהר מהפרדת הנתונים ל-Train ו-Test.
 אם הפרדת הנתונים לא נעשתה נכון, ייתכן שיהיה קושי להבחין בין הנתונים.

הערה: יש להיזהר מהפרדת הנתונים ל-Train ו-Test.

K-fold cross validation

הפרדת הנתונים ל-Train ו-Test
 K-fold cross validation

20%	20%	20%	20%	20%
-----	-----	-----	-----	-----

הפרדת הנתונים ל-Train ו-Test
 K-fold cross validation
 $\binom{5}{4} = 5$

הפרדת הנתונים ל-Train ו-Test

$$\sum_{i=1}^5 \sum_{j=1}^5$$

$$\sum = \frac{1}{5} \sum_{i=1}^5 \sum_{j=1}^5$$

הפרדת הנתונים ל-Train ו-Test