

17 ripe

הצורה הכללית של פונקציית הפסד היא:

הפסד = $\sum_{i=1}^n \text{loss}(y_i, \hat{y}_i)$

כאשר y_i היא התוצאה האמיתית ו- $\hat{y}_i$ היא התוצאה המוערכת.

Gradient Descent - שיטה

הענין שלום המצב הא הא $\mathbb{R}^d$ כולו, $d \geq 1$
 וכן נתון $x \in \mathbb{R}^d$ ו- $r > 0$ נתון, $B(x, r)$ המכונה

$$\vec{w} = (w_1, \dots, w_d) ; \quad \vec{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$$

הערה: יש להבין את המושגים הנ"ל כהקדמה לשיעור הבא.

עודף, כיצד זה יתבטא ב- $R \geq \mu_c$?

עליון (15) בן 2 (2-15) (עליון 15)

ידיע גרענען אפן שטעלע ביי (גרענע אפן יידיש)

$$\phi = \langle \vec{w}, \vec{x} \rangle = \vec{w} \cdot \vec{x}^T = \sum_{i=1}^d w_i \cdot x_i$$

$\phi = \text{np.dot}(x, w)$ dot product $\rightarrow$ inner product

$= \text{np.dot}(x, w)$

שם בן/בת: דניאל בן דוד

א.ד.נ. ו' ד' תשנ"ב, י"ג שבט

למה זה נקרא הספד

$$f(\vec{x}, \vec{w}) = w_0 \cdot 1 + w_1 x_1 + \dots + w_n x_n$$

אני נמצא כאן כדי לסייע לך. w_0 הוא המעמד הראשוני.
 במרחב דו-ממדי $\mathbb{R}^2$ מופיע
 הווקטור w שלפניו מיוחס

סך הריבועים

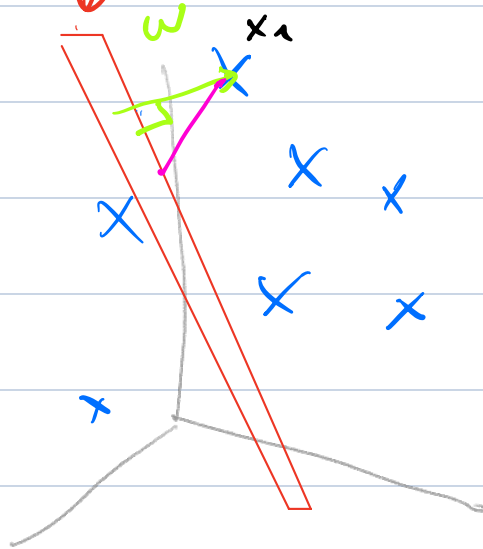
$$\# J = \sum_{i=1}^n (\hat{y}_i - y_i)^2 = \sum_{i=1}^n [(\vec{w} \cdot \vec{x}_i) - y_i]^2$$

$\vec{y}_i = \vec{w} \cdot \vec{x}_i$

כאן הדינאמיקה אופיינית היא x_i ויכול להיות w במקום.

שימו לב: בדרך כלל הדינאמיקה x_i מתנהגת $\vec{w}$ היא משתנה
 שזה אנו מחפשים להעריך הקבוצה $\vec{w}$ שבה
 שיהיה מינימלי.

הערה: בניגוד ל $\vec{w}$ מהתנהגות בלתי-משתנה היא תלויה
 במידה.



הערה: אין מילוי מחדש הקובץ כדי קרובה למילוי
 זה יש קרובה שבמקום לא-מילוי מילוי

והייתם אלו את המודל של הנתונים.

אם יש לנו נתונים $\{x_i, y_i\}_{i=1}^n$ ורצונו למצוא את המודל $\vec{w}$ ואת הבias b כך שיהיה:

הפונקציה $f(x) = \vec{w} \cdot \vec{x} + b$ תתאים טוב לנתונים. (הערות: $\vec{w}$ ו- b הם וקטורים).

אם $\vec{w}$ ו- b הם וקטורים, אז $\vec{w} \cdot \vec{x} + b$ הוא סכום של n איברים. האם יש דרך למצוא את $\vec{w}$ ו- b ?

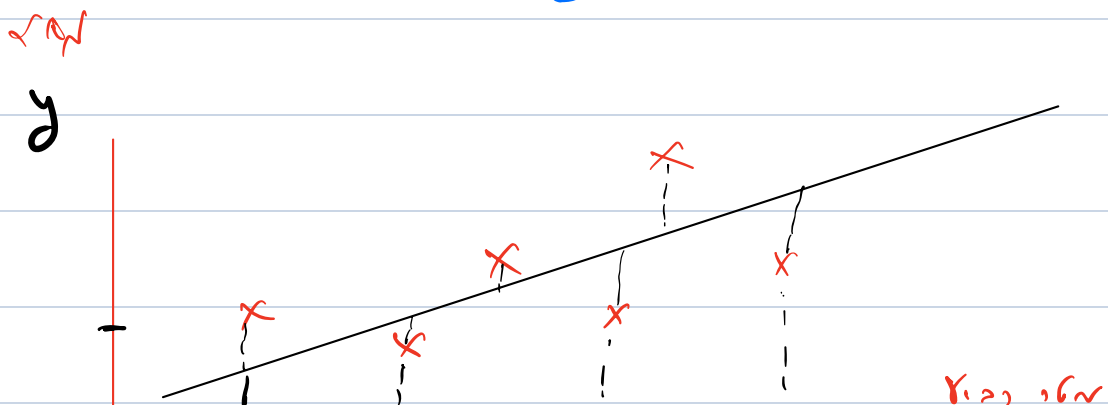
כן, $O(n)$.

אם $\vec{w}$ ו- b הם וקטורים, אז $\vec{w} \cdot \vec{x} + b$ הוא סכום של n איברים. האם יש דרך למצוא את $\vec{w}$ ו- b ?

$$J = \sum_{i=1}^n (\hat{y}_i - y_i)^2 = \sum_{i=1}^n [(\vec{w} \cdot \vec{x}_i) - y_i]^2$$

הערות:
הערות: $\vec{w}$ ו- b הם וקטורים.

אנחנו רוצים למצוא את $\vec{w}$ ו- b כך שיהיה:



הערכה: באמצעות המידע שלנו, $x \in \mathbb{R}^d$ למשל,
 נחזיק המטרה היא למצוא y_i ולכן כנסים
 (המטרה היא $d+1$).

$$\{x_i, y_i\}$$

כדי לקבל מניחה J , עלינו למצוא פונקציה J למשל:

I. פונקציה קוויadrטית

II. פונקציה קוויadrטית.

$$\nabla J = \frac{\partial J}{\partial \vec{w}} = 2 \sum_{i=1}^n (\vec{w} \cdot \vec{x}_i - y_i) \cdot \vec{x}_i$$

על מנת שיהיה זה המינימום של J , עלינו למצוא את הנקודה שבה $\nabla J = 0$.

האלגוריתם הזה נקרא Gradient Descent:

$$\vec{w}^{(t+1)} = \vec{w}^{(t)} - \alpha \cdot 2 \sum_{i=1}^n (\vec{w}^{(t)} \cdot \vec{x}_i - y_i) \cdot \vec{x}_i$$

הכיוון של $\vec{w}^{(t+1)}$ הוא הפוך לזה של $\vec{w}^{(t)}$ (הוא נע במידת האפשרות).

כל הפעולה היא למצוא את המינימום של J .

הערה: α הוא קבוע $0 < \alpha < 1$ שנקרא Learning Rate.

אנחנו, שלא בדיוק הנדסים! ראה עיניך ו
 תירוב אנחנו

קצת עבתי האמלגם המטריציוני

$$\textcircled{\text{☺}} \quad \frac{\partial J}{\partial w_j} = 2 \sum_{i=1}^n (\hat{y}_i - y_i) \cdot x_{i,j}$$

נכני עיניך

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \stackrel{\text{הכנסה}}{=} \begin{pmatrix} x_{1,1} & \dots & x_{1,d+1} \\ \vdots & & \vdots \\ x_{n,1} & \dots & x_{n,d+1} \end{pmatrix}$$

$$X \in M^{n \times (d+1)}$$

עברתי
 מה עברתי $\vec{w}$ כקטע

$$w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{d+1} \end{pmatrix} \in M^{1 \times (d+1)}$$

אנחנו לא צריכים "עברתי" כלל הנהגה

כאן נכני

$$\sum_{i=1}^n (\vec{w} \cdot \vec{x}_i^T) = \sum_{i=1}^n (w_1 \dots w_{d+1}) \begin{pmatrix} x_{i,1} \dots \\ x_{i,d+1} \end{pmatrix} \cdot x_{i,j}$$

הצגת המטריצה X כמטריצה $n \times (d+1)$

$$\begin{aligned}
 &= \sum_{i=1}^n \sum_{k=1}^{d+1} (w_k \cdot x_{i,k}) \cdot x_{i,j} \\
 &= \sum_{i=1}^n \sum_{k=1}^{d+1} w_k \cdot x_{i,k} \cdot x_{i,j} \\
 &= \sum_{i=1}^n \sum_{k=1}^{d+1} x_{i,j} \cdot x_{i,k} \cdot w_k \\
 &= (X^T \cdot X \cdot w)_j
 \end{aligned}$$

הצגת y כמטריצה $n \times 1$

$$\sum_{i=1}^n y_i \cdot x_{i,j} = \sum_{i=1}^n x_{j,i}^T \cdot y_i = (X^T \cdot y)_j$$

הצגת y כמטריצה $n \times 1$

הצגת w כמטריצה $(d+1) \times 1$

הצגת w כמטריצה $(d+1) \times 1$

$$\frac{\partial J}{\partial w_j} = 2 \sum_{i=1}^n (\hat{y}_i - y_i) \cdot x_{i,j}$$

הצגת w כמטריצה $(d+1) \times 1$

$$\frac{\partial J}{\partial w} = 2 [X^T \cdot X \cdot w - X^T \cdot y] = 2 X^T [X \cdot w - y]$$

הצגת w כמטריצה $(d+1) \times 1$

$$w^{(t+1)} - w^{(t)} = \alpha \left[\underbrace{x^T \cdot x \cdot w^{(t)}}_{\text{מכפלה}} - \underbrace{x^T \cdot y}_{\text{מכפלה}} \right]$$

הגדלה

כאשר המהלך של המכפלה $x^T \cdot x \cdot w^{(t)}$ הוא זהה לזה של המכפלה $x^T \cdot y$.

נניח כי

$$\frac{\partial J}{\partial w} = x^T \cdot x \cdot w - x^T y = 0$$

$$x^T \cdot x \cdot w = x^T \cdot y$$

||

$$(x^T \cdot x) \cdot w = x^T \cdot y$$

נכפול במעריך המהפכי $(x^T \cdot x)^{-1}$.

$$w = (x^T \cdot x)^{-1} \cdot x^T \cdot y$$

הנחה / הערה

$$x^T \cdot x \text{ המכפלה}$$

$$x^T \cdot x \text{ המכפלה}$$

היא תהיה כזו:

$$x = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$$

היא תהיה כזו:

נרשם לרשימה מסתמך על סיווג, מסתמך על
מצב למעשה על דיון רחב יותר - סיווג

רשימה לרשימה

רשימה לרשימה - רשימה של רשימה (רשימה)
בין, כולל יש רק אחד מהם. רשימה
כך/כך לא.

בשם הרשימה הרשימה רשימה רשימה רשימה
רשימה רשימה רשימה רשימה רשימה רשימה
 $\{(x_i, y_i)\}_{i=1}^n$ רשימה רשימה רשימה
רשימה רשימה רשימה רשימה רשימה רשימה
train data?

רשימה רשימה רשימה רשימה רשימה רשימה

$$\hat{y} = \begin{cases} 0 & \vec{w} \cdot \vec{x} < 0 \\ 1 & \vec{w} \cdot \vec{x} \geq 0 \end{cases}$$

רשימה רשימה רשימה רשימה רשימה רשימה

רשימה רשימה רשימה רשימה רשימה רשימה

$$\vec{x} \rightarrow \vec{w} \cdot \vec{x} \xrightarrow{g(\cdot)} \hat{y}$$

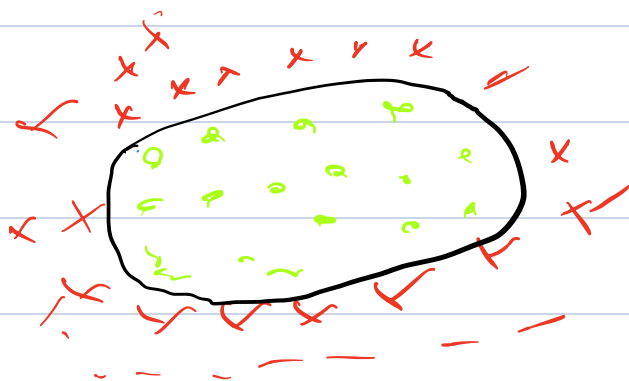
רשימה רשימה רשימה רשימה רשימה רשימה

רשימה רשימה רשימה רשימה רשימה רשימה

$$g(z) = \begin{cases} 1 & z \geq 0 \\ 0 & z < 0 \end{cases}$$

כאן בחרנו לעבור ההכרזה הזו לפרק.

ישו לב: "אם ורק אם" אלו מונחים יחד, נחלק (הכרזה לא פתוחה, אלא קבוצה היא).



בטור 15 נבין איך זה ייתכן (kernel trick).

המורה שלנו אמן ממש מוכשר לעצמה
מוקדש - מצטט, ציף ציפה. מביט?

(1) בקדו - המוקדש ביהוה

SK המוקדש ציפה לקבל - היות 1

על מ $z \rightarrow -\infty$ המוקדש ציפה לקבל 0

(2) באמצע המוקדש ציפה לקבל רצף

תדפיסה: יש לנו פונקציה ששכיחה את התוצאה 32

הינה שם האפוסטרוף הוא שם 1 כי

הפונקציה של המונה והמכנה ציין אותה נראה צורה

רצוי לדבר על זה, יש כמה פונקציות ששכיחה

(הינה שם צורה הפונקציה ששכיחה המונה)

על זה? e^x יכולה להיות גם לא

$$\lim_{x \rightarrow -\infty} e^x = \lim_{x \rightarrow +\infty} \frac{1}{e^x} = 0$$

כמו שאמרו, נוסח המונה ששכיחה שם

אין קשר כלל וכלל (בטח מהן $\pm$ ואין $\neq$)

משפט אחר:

(3) הינה ששכיחה אותה ציין אותה שם האפוסטרוף

אפוסטרוף, כנראה שם כן נראה בטח סקד

משפט נראה כזה, רצוי ונראה

$$\hat{y} = w_0 + w_1 \cdot x + \epsilon$$

כאשר ϵ הוא אפוסטרוף ששכיחה

ובאופן כללי נראה $\vec{z} = \vec{w} \cdot \vec{x}$

הינה למעשה רצוי הוא $z \in \mathbb{R}$

ואם נראה שם כן נראה פונקציה

אנחנו

$$\hat{y}_i = \frac{1}{1 + e^{-\vec{w} \cdot x_i}}$$

כל הבעיה

הבעיה היא להחליט על $\vec{w}$ כדי שיהיה
כל $y_i \in \{0, 1\}$ ו- $0 \leq y_i \leq 1$

כל הבעיה