27/3/20 - ו׳

תרגול 2

<u>הנושאים</u>: 1. Level Sets
2. נגזרת
3. גרדיאנט
4. הצגה לינארית
5. היעקוביאן
6. ההסיאן

נדבר על פונק׳ מ-$\mathbb{R}^n$ ל-$\mathbb{R}$. למידה מנסה למצע את הפונק׳ הזו.
כלומר נמצא את $\underset{\omega \in \mathbb{R}^n}{\mathrm{argmin}}\, f(\omega)$, ל-$f: \mathbb{R}^n \to \mathbb{R}$.

<u>איך ממצעים פונק׳ סקלרית</u>? אם הפונק׳ חלקה אז גוזרים ומשווים לאפס, נק׳ אלו הן נק׳ הקיצון ואז נסווג לנק׳ מינ׳, מקס׳ או אוכף. מהן נקח את נק׳ המינימום הלוקלית.

★ נעשה את אותו הדבר, רק שהנקודות יהיו וקטורים.

(1) Level Sets

• <u>level sets</u>: של הפונק׳ $f: \mathbb{R}^d \to \mathbb{R}$ עבור ערך מסוים $c \in \mathbb{R}$ היא הקבוצה: $\{x \mid f(x) = c\}$

<u>תרגיל 1</u>: נמצא את ה-level set של הפונק׳: $f(x,y) = x^2 + y^2$:
עבור $c \in \mathbb{R}$: $f(x,y) = x^2+y^2 = c$ מתאר מעגל ברדיוס $\sqrt{c}$ סביב $(0,0)$.
לכן "קווי הגובה", כלומר ה-level sets יהיו כל המעגלים הללו!

(2) נגזרת

• <u>נגזרת</u>: תהא $f: \mathbb{R} \to \mathbb{R}$ פונק׳ סקלרית. הנגזרת של f ב-$x \in \mathbb{R}$ היא:
$$\frac{df(x)}{dx} = \lim_{a \to 0} \frac{f(x+a) - f(x)}{a}$$

<u>דוגמא</u>: פונק׳ ה-ReLU: $f(x) = x^+ = \max(0,x)$. <u>אזי</u>: $\frac{df(x)}{dx} = \begin{cases} 0 & : x<0 \\ 1 & : x>0 \end{cases}$

③ גרדיאנט

• **נגזרת חלקית**: תהא $f: \mathbb{R}^n \to \mathbb{R}$. הנגזרת החלקית של f ב-$x \in \mathbb{R}^n$ ביחס לרכיב ה-i, x_i, היא:

$$\frac{\partial f(x)}{\partial x_i} = \lim_{a \to 0} \frac{f(x+ae_i) - f(x)}{a}$$

דוגמאות: • אזי: $f(x) = \max_i (x_1, .., x_n)$,

$$\frac{\partial f(x)}{\partial x_i} = \begin{cases} 0 & \text{אחרת:} \\ 1 & i = \arg\max (x_1, .., x_n): \end{cases}$$

• $f(x,y) = x^2 + xy + y^2$. אזי:

$$\frac{\partial f(x_0,y_0)}{\partial x} = 2x_0 + y_0$$

$$\frac{\partial f(x_0,y_0)}{\partial y} = 2y_0 + x_0 \quad \left(\Longrightarrow \quad \nabla f(x_0,y_0) = (2x_0+y_0,\ 2y_0+x_0)^T \right)$$

• **גרדיאנט**: הגרדיאנט של $f: \mathbb{R}^n \to \mathbb{R}$ ב-$x \in \mathbb{R}^n$ הוא וקטור הנגזרות החלקיות:

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix} \in \mathbb{R}^n$$

(זה החץ הקטן שמוטל על המישור xy בהטלה של x. הם מצביעים ממינימום לוקלי למקסימום לוקלי, ותמיד הולכים מעלה. המשמעות שהם הוא: הכיוון מצביע מעלה-הכיוון בו הפונק' עולה הכי הרבה מהנק' הנ"ל, והגודל מראה כמה חזק הפונק' עולה.)

• **דוג'**: $f(x) = b^T x = \sum_{i=1}^{d} b_i x_i$ עבור $f: \mathbb{R}^d \to \mathbb{R}$, $b \in \mathbb{R}^d$. אזי:

$$\frac{\partial f}{\partial x_j} = b_j \Longrightarrow \nabla f(x) = \begin{pmatrix} b_1 \\ \vdots \\ b_d \end{pmatrix} = b$$

• $f: \mathbb{R}^d \to \mathbb{R}$, $f(x) = \|x\|^2 = \sum_{i=1}^{d} x_i^2$. אזי: $\frac{\partial f}{\partial x_j} = 2x_j \Longrightarrow \nabla f(x) = 2x$

טענה (תנאי הכרחי [לא מספיק] לנק' קיצון של פונק'): תהא $f: \mathbb{R}^n \to \mathbb{R}$. אם x_0 הוא נק' קיצון של f ו-$\forall i$: $\frac{\partial f(x_0)}{x_i}$ קיימת, אזי: $\nabla f(x_0) = 0$.

④ קירוב לינארי

טור טיילור של הפונק' הסקלרית $f: \mathbb{R} \to \mathbb{R}$ הוא:

$$T(x_0+x) = \underbrace{\sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}}_{\text{שגזרה } \infty \text{ פעמים, סביב הנק' } x_0} x^n$$

☆ נבחין שבה טור. מצטמצם ב-x. אם $x \ll 1$ אזי חזקות x הולכות וקטנות ואפשר להסתפק בסדרים הנמוכים של x. הסדר המוביל הוא הלינארי (כאשר x בחזקת 1), ולכן נסתפק רק ב-2 איברי הטור הראשונים:

$$f(x_0+x) \approx f(x_0) + f'(x_0) \cdot x.$$

קירוב ליניארי: (הכללת טור טיילור למקרה רב-מימד.): תהא $f: \mathbb{R}^n \to \mathbb{R}$ ו- $p_0 \in \mathbb{R}^n$. הקירוב הליניארי של f ב- p_0: $f(p_0) + \langle \nabla f(p_0), (p-p_0) \rangle$

כלומר שלכל הסחה p של p_0: $f(p_0+p) \approx f(p_0) + \langle \nabla f(p_0), p \rangle$

דוגמ: • $f: \mathbb{R}^2 \to \mathbb{R}$, $p_0 = (x_0, y_0)$:

$$f(x_0+x, y_0+y) \approx f(x_0,y_0) + x \cdot \frac{\partial f(x_0,y_0)}{\partial x} + y \cdot \frac{\partial f(x_0,y_0)}{\partial y}$$

• $b \in \mathbb{R}^d$, $f: \mathbb{R}^d \to \mathbb{R}$ כאשר: $f(x) = b^T x = \sum_{i=1}^{d} b_i x_i$, אזי:

$$f(p_0) + \langle \nabla f(p_0), (p-p_0) \rangle = b^T p_0 + \langle b, (p-p_0) \rangle = b^T (p_0 + (p-p_0)) = f(p)$$

★ כלומר שהקירוב הליניארי של פונק' ליניארית הוא הפונק' עצמה (כמובן..)

• נמצא את הקירוב הליניארי של $f(x,y) = \sqrt{x^2+y^2}$ ב: $p_0 = (3,4)$:

(נסמן $p = (x,y)$ אזי: $f(3+x, 4+y) \approx 5 + \underbrace{\frac{3}{5}x + \frac{4}{5}y}_{(*)}$

$$(*)\quad \nabla f(3,4) = \left(\frac{x_0}{\sqrt{x_0^2+y_0^2}}, \frac{y_0}{\sqrt{x_0^2+y_0^2}} \right)^T (3,4) = \left(\frac{3}{5}, \frac{4}{5} \right)^T$$

•

מה ★ הגרדיאנט מורה על הכיוון שבו נמקסם את ערך הפונק' f. למה?

$f(p_0+p) \approx f(p_0) + \nabla f(p_0) \cdot p$, כאשר: $\nabla f(p_0) \cdot p = \|\nabla f(p_0)\|_2 \|p\|_2 \cos\theta$,

θ הזווית בין: $\nabla f(p_0), p$. $\cos\theta \in [-1,1]$, אזי הביטוי מקס' כאשר $\cos\theta = 1$,

כלומר כאשר $\theta = 0°$. הביטוי מינ' כאשר $\cos\theta = -1$, ואפס כאשר $\theta = 90°$. אזי

הכיוון הנה הוא: $p = \frac{-\nabla f(p_0)}{\|\nabla f(p_0)\|_2}$, והכיוון שממקסם: $p = \frac{\nabla f(p_0)}{\|\nabla f(p_0)\|_2}$

⑤ היעקוביאן

נזכר על פונק' $f: \mathbb{R}^d \to \mathbb{R}^m$, הן $\overbrace{}^{\text{מהצורה}}$: $f(x) = (f_1(x), \dots, f_m(x))$ ל- $x \in \mathbb{R}^d$

כאשר $f_i: \mathbb{R}^d \to \mathbb{R}$ לכל $i \in [m]$.

היעקוביאן: תהא $f: \mathbb{R}^d \to \mathbb{R}^m$. היעקוביאן הוא $J_x(f) \in M_{m \times d}$ בנק' x:

(הכללה נוספת של מושג הנגזרת)

$$J_x(f) = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \cdots & \frac{\partial f_1(x)}{\partial x_d} \\ \vdots & & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \cdots & \frac{\partial f_m(x)}{\partial x_d} \end{bmatrix} = \begin{bmatrix} \vdash \nabla f_1(x)^T \dashv \\ \vdots \\ \vdash \nabla f_m(x)^T \dashv \end{bmatrix}$$

דוגמ: • $f: \mathbb{R}^2 \to \mathbb{R}$, $f(x) = x_1^2 + x_2^2$. אזי: $\frac{\partial f}{\partial x_1} = 2x_1$, $\frac{\partial f}{\partial x_2} = 2x_2$

ולכן: $J_x = (2x_1, 2x_2)$, $\nabla f(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$

• $A\in\mathbb{R}^{m\times n}$, $f:\mathbb{R}^n\to\mathbb{R}^m$, $f(x)=Ax$. אזי: $f(x)=\begin{pmatrix}f_1(x)\\ \vdots\\ f_m(x)\end{pmatrix}$

בנוסף: $f_i(x)=\sum_{j=1}^{n}a_{ij}x_j$ ל- $i\in[m]$, ולכן

כמו כן: $\frac{\partial f_i}{\partial x_j}=a_{ij}$, ולכן: $J_x(f)=A$.

כלל השרשרת: יהיו $f:\mathbb{R}^d\to\mathbb{R}^m$, $g:\mathbb{R}^k\to\mathbb{R}^d$, ונקבל כי: $f\circ g:\mathbb{R}^k\to\mathbb{R}^m$.

אז היעקוביאן של $f\circ g$ בנק׳ $x\in\mathbb{R}^k$ הוא:

$$J_x(f\circ g)=J_{g(x)}(f)J_x(g)=\begin{bmatrix}\frac{\partial f_1(g(x))}{\partial g_1(x)} & \cdots & \frac{\partial f_1(g(x))}{\partial g_d(x)}\\ \vdots & & \\ \frac{\partial f_m(g(x))}{\partial g_1(x)} & \cdots & \frac{\partial f_m(g(x))}{\partial g_d(x)}\end{bmatrix}\begin{bmatrix}\frac{\partial g_1(x)}{\partial x_1} & \cdots & \frac{\partial g_1(x)}{\partial x_k}\\ \vdots & & \\ \frac{\partial g_d(x)}{\partial x_1} & \cdots & \frac{\partial g_d(x)}{\partial x_k}\end{bmatrix}$$

(ובכתיב אחר: $J_x(f\circ g)_{i,j}=\sum_{l}\frac{\partial f_i(g(x))}{\partial g_l(x)}\frac{\partial g_l(x)}{\partial x_j}$)

דוגמה: $g(x)=Ax$, $f(x)=\|x\|^2$. אזי:

$f(g(x))=\|Ax\|^2$

$J_{g(x)}(f)=2x^TA^T$, $J_x(g)=A \Rightarrow J_x(f\circ g)=2x^TA^TA$

• MSE: רוצים למזער את $\frac{1}{2}\|A^Tx-y\|^2$ (כשמזערים על $x\in\mathbb{R}^n$).

איך? נגזור ונשווה ל-0.

$$\frac{1}{2}\|A^Tx-y\|^2\overset{T\text{ עבר}}{=}\frac{1}{2}\left((x^TA-y^T)(A^Tx-y)\right)=\frac{1}{2}x^TAA^Tx-y^TA^Tx+\frac{1}{2}\|y\|^2$$

$$=\frac{1}{2}[x^TAA^Tx-y^TA^Tx-x^TAy+y^Ty]$$

הגרדיאנט של זה (אותו דבר)

נ״ל: ונשווה ל-0: $A(A^Tx-y)$

ש-x הפתרון צריך לקיים את המשוואה: $AA^Tx=Ay$.

• Softmax function: $f:\mathbb{R}^n\to\mathbb{R}^n$ כאשר $y_i=f(x)\in(0,1)$

$$y_j=\frac{e^{a_j}}{\sum_{k=1}^{n}e^{a_k}}$$ ← נרמול

וכן: $\sum_{i=1}^{n}y_i=1$. הוא נותן ע״:

נחזין גם שהמשקל נמצא בסדר עולה. הוא קירוב רך של הפונק׳:

hardmax: $f(x)=\max_{i\in[n]}(x_i)$

נסמן: $h=\sum_{k=1}^{n}e^{a_k}$, $g_i=e^{a_i}$. אזי:

$$\frac{\partial y_i}{\partial a_j}=\frac{\partial}{\partial a_j}\frac{e^{a_i}}{\sum_{k=1}^{n}e^{a_k}}=\frac{\partial}{\partial a_j}\frac{g_i}{h}$$

ונקבל:

(1) $i=j$: $$\frac{\partial}{\partial a_j}\frac{e^{a_i}}{h}=\frac{g_i\cdot h-g_j\cdot g_i}{h^2}=\frac{g_i}{h}\cdot\frac{h-g_i}{h}=$$
$$=y_i\cdot(1-y_j)$$

(2) $i\neq j$: תרגיל.

(6) הסיאן

הסיאן: $f: \mathbb{R}^n \to \mathbb{R}$. הכללה רב-מימדית של הנגזרת ה-2.

מטריצת ההסיאן $H_f \in M_n$ של פונק' f מוגדרת להיות:

$$H_f = \begin{bmatrix} \frac{\partial^2 f}{\partial^2 x_1} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial^2 x_n} \end{bmatrix}$$

דוג': $f(x,y) = x^2 + xy + y^2$ אזי.

$$\frac{\partial f(x,y)}{\partial x} = 2x + y, \quad \frac{\partial f(x,y)}{\partial y} = 2y + x$$

$$\frac{\partial^2 f(x,y)}{\partial^2 x} = 2, \quad \frac{\partial^2 f(x,y)}{\partial^2 y} = 2, \quad \frac{\partial^2 f(x,y)}{\partial x \partial y} = \frac{\partial^2 f(x,y)}{\partial y \partial x} = 1$$

ונקבל:

$$H_f = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

(7) נספח על חישוב ומשמעות ההסיאן

* משמעות (בין היתר): מאפשר להבין אם נק' קיצון היא מקס או מינ' לוקאלי, או אוכף, ע"י בחינת הסימן של הע"ע של H_f.