פתרון מטלה 06 — חשבון אינפיניטסימלי 3, 80415

21 במאי 2025

שאלה 1

נתונה רשת נוירונים מהצורה הבאה

$$\mathbb{R}^2 \xrightarrow{W^1} \mathbb{R}^3 \xrightarrow{\alpha^1} \xrightarrow{W^2} \mathbb{R}^2 \xrightarrow{\alpha^2} \mathbb{R}^2$$

המידע הנחוץ על הרשת הוא

1. $W^1 \in M_{3\times 2}(\mathbb{R}), W^2 \in M_{2\times 3}(\mathbb{R})$
2. $\alpha^1 : \mathbb{R}^3 \to \mathbb{R}^3, \alpha^2 : \mathbb{R}^2 \to \mathbb{R}^2$ פועלות בתור ReLU על כל קורדינאטה בנפרד, כאשר $\text{ReLU} : \mathbb{R} \to \mathbb{R}$ מוגדרת באמצעות $\text{ReLU}(t) = \max\{t, 0\}$ וניתן להתעלם מהאי גזירות של ReLU ב־0
3. פונקציית ההפסד $C : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ היא $C(x,y) = \|x-y\|_2^2$

נניח כי נתון לנו זוג קלט פלט $(x,y) \in \mathbb{R}^2 \times \mathbb{R}^2$ ונגדיר את הוקטורים

$$x^0 = x,\ x^1 = \alpha^1(W^1x^0),\ x^2 = \alpha^2(W^2x^1)$$

ונגדיר את הפונקציות הבאות ל־$\mathbb{R}$:

1. $g^2(z^2) = C(\alpha^2(z^2), y)$ עבור $z^2 \in \mathbb{R}^2$
2. $h^2(W^2) = g^2(W^2x^1)$ עבור $W^2 \in M_{2\times 3}(\mathbb{R})$
3. $g^1(z^1) = g^2(W^2\alpha^1(z^1))$ עבור $z^1 \in \mathbb{R}^3$
4. $h^1(W^1) = g^1(W^1x^0)$ עבור $W^1 \in M_{3\times 2}(\mathbb{R})$

ונתון

$$x = \begin{pmatrix}1\\2\end{pmatrix},\ y = \begin{pmatrix}0\\0\end{pmatrix},\ W^1 = \begin{pmatrix}3 & -1\\-2 & 0\\0 & 1\end{pmatrix},\ W^2 = \begin{pmatrix}0 & -1 & 1\\-1 & 4 & 1\end{pmatrix}$$

סעיף א'

נחשב את $\delta^2 = \nabla g^2(W^2x^1)$ ואת $\nabla h^2(W^2)$.

פתרון: ראשית ניזכר

$$(\star)\,\text{ReLU}'(t) = \begin{cases}0 & t<0\\ \times & t=0\\ 1 & t>0\end{cases}$$

ועכשיו נחשב

$$z^1 = W^1x = \begin{pmatrix}3 & -1\\-2 & 0\\0 & 1\end{pmatrix}\begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}3-2\\-2\\2\end{pmatrix} = \begin{pmatrix}1\\-2\\2\end{pmatrix}$$

$$x^1 = \alpha^1(z^1) = \text{ReLU}(z^1) = \begin{pmatrix}\max(1,0)\\\max(-2,0)\\\max(2,0)\end{pmatrix} = \begin{pmatrix}1\\0\\2\end{pmatrix}$$

$$z^2 = W^2x^1 = \begin{pmatrix}0 & -1 & 1\\-1 & 4 & 1\end{pmatrix}\begin{pmatrix}1\\0\\2\end{pmatrix} = \begin{pmatrix}2\\1\end{pmatrix}$$

$$x^2 = \alpha 2(z^2) = \text{ReLU}(z^2) = \begin{pmatrix}2\\1\end{pmatrix}$$

עבור פונקציית ההפסד מתקיים

$$g^2(z^2) = C(\text{ReLU}(z^2), y) = \left\|\text{ReLU}(z^2) - y^2\right\|_2^2 \underset{y=\binom{0}{0}}{=} \left\|\text{ReLU}(z^2)\right\|^2$$

ואז עבור הנגזרת מכיוון ש-$y=\binom{0}{0}$ מתקיים

$$\nabla C(x^2,y)=2x^2=2\begin{pmatrix}1\\2\end{pmatrix}=\begin{pmatrix}4\\2\end{pmatrix}$$

אבל $x^2=\mathrm{ReLU}(z^2)$, אז נשתמש בכלל השרשרת ונקבל

$$\delta^2=\nabla g^2(z^2)=2\big((x^2-y)\odot\mathrm{ReLU}'(z^2)\big)\underset{(\star)}{=}\begin{pmatrix}4\\2\end{pmatrix}\odot\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}4\\2\end{pmatrix}$$

עכשיו

$$\nabla h^2(W^2)_{ij}=\delta_i^2\cdot x_j^1\Rightarrow\nabla h(W^2)=\delta^2\cdot {x^1}^t=\begin{pmatrix}4\\2\end{pmatrix}(1,0,2)=\begin{pmatrix}4&0&8\\2&0&4\end{pmatrix}$$

□

סעיף ב'

נחשב את $\delta^1=\nabla g^1(W^1x^0)$ ואת $\nabla h^1(W^1)$.

פתרון: נשים לב שמתקיים

$$\left(W^2\right)^t\delta^2=\begin{pmatrix}0&1\\-1&4\\1&1\end{pmatrix}\begin{pmatrix}4\\2\end{pmatrix}=\begin{pmatrix}2\\4\\6\end{pmatrix}$$

ומהגדרת $\mathrm{ReLU}'(t)$ עם סעיף א' מתקיים

$$(D\alpha^1)z^1=D\,\mathrm{ReLU}\left(\begin{pmatrix}1\\-2\\2\end{pmatrix}\right)=\begin{pmatrix}1&0&0\\0&0&0\\0&0&1\end{pmatrix}$$

ולכן

$$\delta^1=(D\alpha^1)z^1\cdot\left(W^2\right)^t\delta^2=\begin{pmatrix}1&0&0\\0&0&0\\0&0&1\end{pmatrix}\begin{pmatrix}2\\4\\6\end{pmatrix}=\begin{pmatrix}2\\0\\6\end{pmatrix}$$

ואז כמו בסעיף א'

$$\nabla h^1(W^1)=\delta^1\cdot\left(x^0\right)^t=\begin{pmatrix}2\\0\\6\end{pmatrix}(1,2)=\begin{pmatrix}2&4\\0&0\\6&12\end{pmatrix}$$

□

שאלה 2

תהיי $A \subseteq \mathbb{R}^k$ ו־A $\gamma : [0,1] \to A$ מסילה הגזירה ב־$(0,1)$ כך ש־$\gamma(0) = \gamma(1)$ (מסילה סגורה).

סעיף א'

נוכיח כי לכל $v \in \mathbb{R}^k$ קיים $t \in (0,1)$ כך ש־v אורתוגונלי ל־$\gamma'(t)$.

הוכחה: יהי $v \in \mathbb{R}^k$ ונגדיר $g : [0,1] \to \mathbb{R}$ על־ידי $g(t) = \langle \gamma(t), v \rangle$.

נגדיר $h(x) = \langle x, v \rangle$ זו פונקציה לינארית ב־x ולכן גזירה בכל מקום, ומתקיים $g(t) = h(\gamma(t))$.

g גזירה בקטע $(0,1)$ כהרכבה של פונקציות גזירות (ורציפה בקטע $[0,1]$ כהרכבה של פונקציות רציפות).

מהנתון שמתקיים $\gamma(0) = \gamma(1)$ נקבל שבפרט מתקיים $g(0) = g(1)$ מהגדרת המכפלה הפנימית.

ממשפט רול קיימת $t \in (0,1)$ כך ש־$g'(t) = 0$, ולכן מכלל השרשרת מתקיים

$$0 = g'(t) = Dh_{\gamma(t)} \circ \gamma'(t) = \langle \gamma'(t), v \rangle$$

ולכן עבור ה־t שמשפט רול מביא לנו נקבל ש־$\gamma'(t)$ ו־v אורתוגונליים.

$\square$

סעיף ב'

נוכיח כי קיים $t \in (0,1)$ כל ש־$\gamma(t)$ אורתוגונלי ל־$\gamma'(t)$.

הוכחה: במטלה הקודמת הראנו שמתקיים עבור $f, g : A \subseteq \mathbb{R}^k \to \mathbb{R}^m$ מתקיים ש־$\langle f, g \rangle : A \to \mathbb{R}$ היא גזירה עם נגזרת הנתונה על־ידי

$$D(\langle f, g \rangle)_x = \langle Df_x, g(x) \rangle + \langle f(x), Dg_x \rangle$$

ולכן נגדיר $g(t) = \langle \gamma(t), \gamma(t) \rangle = \|\gamma(t)\|^2$.

כמקודם, g גזירה בקטע$(0,1)$ כהרכבה של פונקציות גזירות (במטלה הקודמת ראינו גם שהנורמה גזירה בכל נקודה) ורציפה בקטע$[0,1]$ כהרכבת פונקציות רציפות.

מתקיים כל תנאי משפט רול ולכן קיימת $t \in (0,1)$ כך שיתקיים

$$0 = g'(t) = 2\langle \gamma'(t), \gamma(t) \rangle$$

כשהנגזרת חושבה מהתרגיל הקודם + כלל שרשרת (ותכונות המכפלה הפנימית ללינאריות).

בנקודה t שמשפט רול מביא לנו נקבל ש־$\gamma'(t)$ ו־$\gamma(t)$ אורתוגונליים. $\square$

סעיף ג'

נוכיח כי לכל פונקציה גזירה $f : A \to \mathbb{R}$ קיים $t \in (0,1)$ כך ש־$\nabla f(\gamma(t))$ אורתוגונלי ל־$\gamma'(t)$.

הוכחה: נגדיר $g(t) = f(\gamma(t))$ וכמקודם זו הרכבה של פונקציות גזירות בקטע $(0,1)$ ורציפות בקטע $[0,1]$ ולכן גזירה ורציפה בקטעים האלו בהתאמה.

נשים לב שמתקיים

$$g(0) = f(\gamma(0)), g(1) = f(\gamma(1))$$

אבל γ מסילה סגורה ולכן $g(1) - g(0) = 0$, ושוב מהפעלה של משפט ערך הממוצע ומשפט רול נקבל שקיימת $t \in (0,1)$ כך שמתקיים $g'(t) = 0$, אבל לפי כלל השרשרת מתקיים

$$g'(t) = \nabla f(\gamma(t)) \cdot \gamma'(t)$$

ומתקיים $\nabla f(\gamma(t)) \cdot \gamma'(t) = 0$, אבל מכפלה סקלרית 0 דהיינו שהם אורתוגונליים, משמע ∇f ו־$\gamma'(t)$ אורתחוגונליים. $\square$

סעיף ד'

נסיק כי אם $f(\gamma(t)) = c$ לכל $t \in [0,1]$ עבור $c \in \mathbb{R}$ קבוע אז $\nabla f(\gamma(t))$ אורתוגונלי ל־$\gamma'(t)$ לכל $t \in (0,1)$.

במילים אחרות, אם התמונה של γ מוכלת בקו־הגובה של $f^{-1}\{c\}$ אז $\nabla f(\gamma(t))$ אורתוגונלי לישר המשיק ל־$\gamma(t)$.

הוכחה: נשתמש ב־g מהסעיף הקודם ומתקיים $g(t) = c$ עבור $c \in \mathbb{R}$ קבוע. נגזור ולפי הסעיף הקודם נקבל ונקבל

$$g'(t) = \nabla f(\gamma(t)) \cdot \gamma'(t) = 0 = c'$$

וזה בידיוק אומר שלכל $t \in (0,1)$ מתקיים ש־$\nabla f(\gamma(t))$ אורתוגונלי ל־$\gamma'(t)$. $\square$

שאלה 3

יהיו $A, B \subseteq \mathbb{R}^k$ פתוחות ו־$f : A \to B$ פונקציה הפיכה כך ש־f גזירה ב־$x \in A$ ו־f^{-1} גזירה ב־$f(x) \in B$.
נוכיח כי Df_x הפיכה עם הופכית הנתונה על־ידי

$$\left(Df_x\right)^{-1} = \left(Df^{-1}\right)_{f(x)}$$

הוכחה: ראשית, f הפיכה ולכן לכל $v \in A$ מתקיים $f \circ f^{-1} = f^{-1} \circ f = I_k$ (כי $Df_x \in M_{k \times k}$).
נעבוד עם כלל השרשרת (כל התנאים ההכרחיים מתקיימים).
ראשית ניזכר שמתקיים $I_k = \left(DI_k\right)_x$, ולכן

$$I_k = \left(DI_k\right)_x = \left(D\left(f^{-1} \circ f\right)\right)_x = \left(Df^{-1}\right)_{f(x)} \circ Df_x$$

$$I_k = \left(DI_k\right)_x = \left(D\left(f \circ f^{-1}\right)\right)_x = Df_x \circ \left(Df^{-1}\right)_{f(x)}$$

משמע מצאנו ש־Df_x הפיכה וההופכית נתונה על־ידי $\left(Df^{-1}\right)_{f(x)}$. $\square$

שאלה 4

תהיי $f:A\subseteq\mathbb{R}^k\to\mathbb{R}^m$ פונקציה.

סעיף א'

נוכיח כי אם f היא M־ליפשיצית וגזירה ב־$x\in A$ אז $\left\|Df_x\right\|_{\text{op}}\le M$.

הוכחה: יהי $v\in\mathbb{R}^k$ וקטור יחידה.

f דיפרנציאבילית ב־$x\in A$ ולכן כל הנגזרות הכיווניות קיימות ובפרט קיימות בכיוון v ומתקיים

$$\|Df_x(v)\|=\|\partial_v f(x)\|=\left\|\lim_{t\to0}\frac{f(x+tv)-f(x)}{t}\right\|\underset{\text{ליפשיציות}}{\le}\lim_{t\to0}\frac{M\|tv\|}{|t|}=\lim_{t\to0}\frac{|t|}{|t|}C\|v\|\underset{\|v\|=1}{=}C$$

ולכן

$$\left\|Df_x\right\|_{\text{op}}=\sup_{\|v\|=1}|Df_x(v)|\le C$$

□

סעיף ב'

נניח כי A היא פתוחה וקמורה, כלומר לכל $x,y\in A$ מתקיים

$$[x,y]:=\{(1-t)x+ty\mid t\in[0,1]\}\subseteq A$$

נוכיח כי אם f גזירה ב־A אז מתקיים

$$\sup\left\{\left\|Df_x\right\|_{\text{op}}\,\middle|\,x\in A\right\}=\inf\{M>0\mid \text{היא } M\text{־ליפשיצית } f\}$$

כאשר המוסכמה היא ש־$\inf\emptyset=\infty$.

הוכחה: מסעיף א' אנחנו מקבלים שמתקיים $\sup_{x\in A}\left\|Df_x\right\|_{\text{op}}\le\inf\{M>0\mid \text{היא } M\text{־ליפשיצית } f\}$ ונראה את ההכלה בכיוון השני:

נגדיר $v=y-x$. מהיות A קמורה ש־$[x,y]\subseteq A$ ונגדיר $h:[0,1]\to\mathbb{R}^n$ על־ידי $f(x+tv)$ שגזירה ורציפה מכך ש־f גזירה ורציפה.

ממשפט ערך הממוצע של הנגזרת נקבל שמתקיים עבור $t_0\in(0,1)$

$$h(1)-h(0)=f(y)-f(x)=h'(t_0)$$

אבל

$$h'(t_0)=\frac{\partial}{\partial t}f(x+tv)=Df_{x+tv}(v)$$

ולכן

$$f(y)-f(x)=Df_{x+t_0v}(v)$$

אבל מתקיים לפי מה שראינו

$$\|f(y)-f(x)\|=Df_{x+t_0v}(v)\le\left\|Df_{x+t_0v}\right\|_{\text{op}}\cdot\|v\|\left\|Df_{x+t_0v}\right\|_{\text{op}}\cdot\|x-y\|$$

ולכן f היא M־ליפשיצית עם $M=\sup_{x\in A}\left\|Df_x\right\|_{\text{op}}$.

קיבלנו הכלה דו־כיוונית ולכן יש שיוויון.

□

שאלה 5

תהיי $f:\mathbb{R}^2\to\mathbb{R}$ הנתונה על־ידי

$$f(x,y)=\begin{cases}\frac{x^3y-xy^3}{x^2+y^2} & (x,y)\neq(0{,}0)\\ 0 & (x,y)=(0{,}0)\end{cases}$$

נוכיח כי הנגזרות החלקיות מסדר שני של f קיימות בראשית אבל f איננה גזירה פעמיים בראשית.

הוכחה: ראשית, נשים לב ש־f רציפה בראשית:

$$\lim_{(x,y)\to(0,0)}\frac{x^3y-xy^3}{x^2+y^2}\underset{\text{קורדינאטות קוטביות}}{=}\lim_{r\to0}\frac{(r\cos(\theta))^3r\sin(\theta)-(r\sin(\theta))^3(r\cos(\theta))}{(r\cos(\theta))^2+(r\sin(\theta))^2}$$
$$\underset{\cos(\theta)^2+\sin(\theta)^2=1}{=}\lim_{r\to0}\frac{r^4\left(\cos(\theta)^3\sin(\theta)-\sin(\theta)^3\cos(\theta)\right)}{r^2}=\lim_{r\to0}r^2\left(\cos(\theta)^3\sin(\theta)-\sin(\theta)^3\cos(\theta)\right)=0$$

נראה שהנגזרות החלקיות מסדר ראשון קיימות:

$$\frac{\partial f}{\partial x}=\frac{-y^5+4x^2y^3+x^4y}{\left(x^2+y^2\right)^2}$$
$$\frac{\partial f}{\partial y}=\frac{x^5-4x^3y^2-xy^4}{\left(x^2+y^2\right)^2}$$

בנקודה $(0,0)$, נחשב את הנגזרות החלקיות ישירות מהגדרה כגבול:

$$\frac{\partial f}{\partial x}(0,0)=\lim_{t\to0}\frac{f(t,0)-f(0,0)}{t}=0$$
$$\frac{\partial f}{\partial y}(0,0)=\lim_{t\to0}\frac{f(0,t)-f(0,0)}{t}=0$$

נראה שהנגזרות החלקיות רציפות ולכן f דיפרנציאבילית פעם אחת (לפחות), נשים לב שהנגזרות החלקיות "סימטריות" עד־כדי הוצאת (-1) גורם משותף (אין משמעות להאם זה x^5 או y^5 לרציפות), ולכן מספיק שנבדוק רציפות של נגזרת חלקית אחת ונקבל מכך גם את של הנגזרת החלקית השנייה:

$$\lim_{(x,y)\to0}\frac{-y^5+4x^2y^3+x^4y}{\left(x^2+y^2\right)^2}\underset{\text{קורדינאטות קוטביות}}{=}=\frac{(-r\sin(\theta))^5+4(r\cos(\theta))^2(r\sin(\theta))^3+(r\cos(\theta))^4(r\sin(\theta))}{\left((r\cos(\theta))^2+(r\sin(\theta))^2\right)^2}$$
$$\underset{\cos(\theta)^2+\sin(\theta)^2=1}{=}\lim_{r\to0}\left(-r^5\right)(\sin(\theta))^5\frac{+}{r^4}=\lim_{r\to0}-r\cos(\theta)^2\left(-1+4\sin(\theta)+\cos^2(\theta)\sin(\theta)\right)=0$$

אז f דיפרנציאבילית בכל נקודה.

עבור הנגזרות השניות נבדוק בראשית ישירות מהגדרת הגבול ונשתמש בגבול לעיל ובנגזרות החלקיות מסדר ראינו שחישבנו בראשית:

$$\frac{\partial^2 f}{\partial x^2}(0,0)=\lim_{t\to0}\frac{\frac{\partial f}{\partial x}(t,0)-\frac{\partial f}{\partial x}(0,0)}{t}=0,\ \ \frac{\partial^2 f}{\partial y^2}(0,0)=\lim_{t\to0}\frac{\frac{\partial f}{\partial y}(0,t)-\frac{\partial f}{\partial y}(0,0)}{t}=0,$$

$$\frac{\partial^2 f}{\partial x\partial y}(0,0)=\lim_{t\to0}\frac{\frac{\partial f}{\partial x}(0,t)-\frac{\partial f}{\partial x}(0,0)}{t}=\lim_{t\to0}\frac{\frac{-t^5}{t^4}}{t}=-1,\ \ \frac{\partial^2 f}{\partial y\partial x}(0,0)=\lim_{t\to0}\frac{\frac{\partial f}{\partial y}(t,0)-\frac{\partial f}{\partial y}(0,0)}{t}=\lim_{t\to0}\frac{\frac{t^5}{t^4}}{t}=1$$

זאת אומרת, הנגזרות החלקיות קיימות אבל $\frac{\partial^2 f}{\partial y\partial x}(0,0)\neq\frac{\partial^2 f}{\partial x\partial y}(0,0)$.

כעת, ניזכר שאם f דיפרנציאבילית פעמיים, הנגזרות החלקיות השניות המעורבות שלה אמורות להיות זהות – שכן פונקציות שדיפרנציבאליות פעמיים מניבות לנו מטריצה סימטרית (למטריצת הנגזרות השניות) (זה נובע גם ממשפט קלרו שהוזכר בתרגול).

לכן הנגזרות החלקיות מסדר שני של f קיימות בראשית אבל f איננה גזירה פעמיים בראשית. □