A. 1. $x \in \mathbb{R}^d$ (וקטור), $f(x) = \|x\|_2^2 = x^T x$ צריך לחשב

אלו הגרדיאנטים של $\nabla f(x)$ - גרדיאנט של נורמת 2!

$$f(x) = \|x\|_2^2 = x^T x = \left(\left(\sum_{i=1}^{d} x_i^2\right)^{1/2}\right)^2 = \sum_{i=1}^{d} x_i^2$$

$$\nabla f(x) = \begin{pmatrix} \frac{\partial x^T x}{\partial x_1} \\ \vdots \\ \frac{\partial x^T x}{\partial x_d} \end{pmatrix}$$

$$\frac{\partial}{\partial x_j} f(x) = \frac{\partial}{\partial x_j} \sum_{i=1}^{d} x_i^2 = \sum_{i=1}^{d} \frac{\partial}{\partial x_j} x_i^2 = \begin{cases} 2x_j & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases}$$

עבור $1 \le j \le d$ ולכן

$$\nabla f_x = \begin{pmatrix} 2x_1 \\ \vdots \\ 2x_d \end{pmatrix} = 2x$$

1. ב. $f(x)=A^T x\in\mathbb{R}^n$, $A\in\mathbb{R}^{d\times n}$ נתון לחשב את
הגרדיאנט של f ביחס ל x. $f:\mathbb{R}^d\to\mathbb{R}^n$

$$f(x)=A^Tx=\begin{pmatrix} a_{11}x_1+a_{21}x_2+\dots a_{d1}x_d \\ a_{12}x_1+a_{22}x_2+\dots a_{d2}x_d \\ \vdots \\ a_{1n}x_1+a_{2n}x_2+\dots a_{dn}x_d \end{pmatrix}=\begin{pmatrix} \sum_{k=1}^{d} a_{k1}x_k \\ \vdots \\ \sum_{k=1}^{d} a_{kn}x_k \end{pmatrix} := \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix}$$

$$\nabla f(x)=\begin{pmatrix} \frac{\partial}{\partial x_1}f_1 & \dots & \frac{\partial f_1}{\partial x_d} \\ \vdots & & \\ \frac{\partial}{\partial x_1}f_n & \dots & \frac{\partial f_n}{\partial x_d} \end{pmatrix}=\begin{pmatrix} \frac{\partial}{\partial x_1}\sum_{k=1}^{d} a_{k1}x_k & \dots & \frac{\partial}{\partial x_d}\sum a_{k1}x_k \\ \vdots & & \\ \frac{\partial}{\partial x_1}\sum a_{kn}x_k & \dots & \frac{\partial}{\partial x_d}\sum a_{kn}x_k \end{pmatrix}\overset{(*)}{=}\begin{pmatrix} a_{11} & a_{21} & \dots a_{d1} \\ a_{12} & a_{22} & a_{d2} \\ \vdots & \vdots & \vdots \\ a_{1n} & a_{2n} & a_{dn} \end{pmatrix}$$

$=A^T\in\mathbb{R}^{n\times d}$

לכל $1\le i\le n$ ולכל $1\le j\le d$ מתקיים: $\frac{\partial}{\partial x_j}f_i=\frac{\partial}{\partial x_j}(a_{1i}x_1+\dots a_{di}x_d)=a_{ji}$

1. ג. $g(x)=A^Tx\in\mathbb{R}^n$, $f(y)=\|y\|_2^2$ נתון לחשב
את הגרדיאנט של $f(g(x))$ ביחס ל x.

$f:\mathbb{R}^n\to\mathbb{R}$

$g:\mathbb{R}^d\to\mathbb{R}^n$

$$f(g(x))=\|A^Tx\|_2^2=(A^Tx)^TA^Tx=x^TAA^Tx:=h(x)$$

נעבור כעת לכלל השרשרת שאינו רכיבה:

$$\boxed{\frac{\partial f(g(x))}{\partial x_i} = \sum_{j=1}^{n} \frac{\partial y_j}{\partial x_i} \cdot \frac{\partial z}{\partial y_j}}$$

במקרה שלנו:

$$y = g(x) = A^T x$$

$$A = \begin{pmatrix} a_{11}, a_{12} \cdots a_{1n} \\ \vdots \\ a_{d1} \cdots \cdots a_{dn} \end{pmatrix} \qquad A^T = \begin{pmatrix} a_{11} & \cdots & a_{d1} \\ a_{12} & & \vdots \\ \vdots & & \vdots \\ a_{1n} & & a_{dn} \end{pmatrix} \in \mathbb{R}^{n \times d}$$

$$A^T x = \begin{pmatrix} a_{11}x_1 + a_{21}x_2 + \cdots a_{d1}x_d \\ a_{12}x_1 + a_{22}x_2 + \cdots a_{d2}x_d \\ \vdots \\ a_{1n}x_1 + a_{2n}x_2 + \cdots a_{dn}x_d \end{pmatrix}$$

$$\frac{\partial y_j}{\partial x_i} = \frac{\partial [A^T x]_j}{\partial x_i} = a_{ij}$$

$$z = f(y) = \|y\|_2^2 = y^T y$$

$$\frac{\partial z}{\partial y_j} = \frac{\partial y^T y}{\partial y_j} = 2y_j$$

$$\Rightarrow \frac{\partial f(g(x))}{\partial x_i} = \sum_{j=1}^{n} a_{ij} 2 y_j = 2\sum_{j=1}^{n} a_{ij} y_j =$$

$$2[a_{i1} y_1 + a_{i2} y_2 + \cdots \cdots a_{in} y_n]$$

$$\Rightarrow \nabla f(g(x)) = 2Ay = \boxed{2 \cdot A A^T x}$$

$$\downarrow$$

$$(A.2\ v) \quad y := A^T x$$

4. a. $g(A) = A^T x \in \mathbb{R}^n$, $f(y) = \|y\|_2^2$, צריך לחשב את הגזירה של $f(g(A))$ מול A.

$g: \mathbb{R}^{d\times n} \to \mathbb{R}^n$, $f: \mathbb{R}^n \to \mathbb{R}$, $f(g(x)) = \|A^T x\|_2^2 = x^T A A^T x$

לפי כלל השרשרת מתקיים

$$\boxed{\frac{\partial}{\partial a_{pq}} f(g(x)) = \sum_{j=1}^{n} \frac{\partial y_j}{\partial a_{pq}} \cdot \frac{\partial z}{\partial y_j}}$$

כאשר $z = f(y) = \|y\|_2^2 = y_1^2 + \dots + y_n^2 = \sum_{i=1}^{n} y_i^2 \Rightarrow \frac{\partial z}{\partial y_j} = 2y_j$

$$y_j = [A^T x]_j = \sum_{k=1}^{d} a_{kj} x_k \Rightarrow \frac{\partial y_j}{\partial a_{pq}} = \frac{\partial}{\partial a_{pq}} \sum_{k=1}^{d} a_{kj} x_k =$$

$$\frac{\partial}{\partial a_{pq}} (a_{1j} x_1 + \dots + a_{dj} x_d) = \begin{cases} x_p & q = j \\ 0 & q \neq j \end{cases} = \mathbb{1}_{(q=j)} \cdot x_p$$

נציב בחזרה בנוסחת השרשרת ונקבל:

$$\frac{\partial}{\partial a_{pq}} f(g(x)) = \sum_{j=1}^{n} \mathbb{1}_{(q=j)} \cdot x_p \cdot 2y_j = x_p \cdot 2y_q$$

נגדיר מטריצות: $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$ מתקיים $AB \in \mathbb{R}^{m\times p}$

$$[AB]_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \quad \text{for } 1 \le i \le m,\ 1 \le j \le p$$

$AB \in \mathbb{R}^{d \times d}$, $B = y^T \in \mathbb{R}^{1 \times d}$, $A = x \in \mathbb{R}^{d \times 1}$ לאחר שלב

ולכן אפק' פ :

$$\nabla f(g(x)) = 2xy^T \underset{y = A^T x}{=} 2x(A^T x)^T = \boxed{2xx^T A}$$

B

6:42 PM Thursday, November 14, 2024

B. $loss = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$

B.1. רשמו את $\hat{y}$ כפונקציה של x, W_1, b_1, W_2, b_2.

סימונים מהבית: $z_1 = W_1 x + b_1$, $a_1 = \frac{1}{1+e^{-z_1}}$

$$\hat{y} = \sigma\left(W_2 \cdot \underbrace{Relu(W_1 x + b_1)}_{a_1} + b_2 + x\right) = \sigma(W_2 \cdot a_1 + b_2)$$

$$= \frac{1}{1+e^{-W_2 a_1 + b_2}} = \frac{1}{1+e^{-(W_2 \cdot Relu(W_1 x + b_1) + b_2 + x)}}$$

B.2. חשבו את $\frac{\partial J}{\partial W_2}, \frac{\partial J}{\partial b_2}$

אופציה ב.

10:52 Sunday, 17 November 2024

ב. $J = -y\log(\hat{y}) - (1-y)\log(1-\hat{y})$

לפי הקודם מגיעים ל $\hat{y} = \sigma(z_2) = Relu(w_2 \cdot a_1 + b_2)$

$$\frac{\partial J}{\partial \hat{y}} = \boxed{\frac{-y}{\hat{y}} + \frac{1-y}{1-\hat{y}}}$$

לפי נגזרת של לוג (מקווה שלא היה צריך לפרט יותר מכך).

$$\frac{\partial \hat{y}}{\partial z_2} = \frac{\partial}{\partial z_2}\left(\frac{1}{1+e^{-z_2}}\right) = \frac{e^{-z_2}}{(1+e^{-z_2})^2} = \frac{1}{1+e^{-z_2}} \cdot \frac{e^{-z_2}}{1+e^{-z_2}} = \boxed{\hat{y}(1-\hat{y})}$$

$$\frac{\partial J}{\partial z_2} = \frac{\partial J}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_2} = \left(\frac{-y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\right) \cdot \hat{y}(1-\hat{y})$$

מתוך השרשרת

$$\frac{\partial z_2}{\partial w_2} = \frac{\partial (w_2 \cdot Relu(z_1) + b_2)}{\partial w_2} = Relu(z_1) = \boxed{\max(z_1, 0)}$$

$$\frac{\partial J}{\partial w_2} = \frac{\partial z_2}{\partial w_2} \cdot \frac{\partial \hat{y}}{\partial z_2} \cdot \frac{\partial J}{\partial \hat{y}} = \max(z_1,0) \cdot \underbrace{\hat{y}(1-\hat{y})\left(\frac{-y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\right)} =$$

מתוך השרשרת

$$\max(z_1, 0)(\hat{y} - y)$$

$$\frac{\partial J}{\partial b_2} = \frac{\partial J}{\partial z_2} \cdot \frac{\partial z_2}{\partial b_2}$$

מתקן שרשרת

$$\frac{\partial z_2}{\partial b_2} = \frac{\partial}{\partial b_2} w_2 \cdot \mathrm{Relu}(z_1) + b_1 = 1$$

$$\frac{\partial J}{\partial b_2} = \frac{\partial J}{\partial z_2} = \left(\frac{-y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\right) \hat{y} \cdot (1-\hat{y}) =$$

מסוף קודם

$$= \frac{-y(1-\hat{y}) + (1-y)\cdot\hat{y}}{\hat{y}\,\cancel{(1-\hat{y})}} \cdot \hat{y}\,\cancel{(1-\hat{y})} =$$

$$-y + \cancel{y\hat{y}} + \hat{y} - \cancel{y\hat{y}} = \hat{y} - y$$

$$\frac{\partial J}{\partial w_1} = \frac{\partial J}{\partial z_2} \cdot \frac{\partial z_2}{\partial a_1} \cdot \frac{\partial a_1}{\partial z_1} \cdot \frac{\partial z_1}{\partial w_1}$$ .3.B

$z_1 = w_1 x + b_1, \quad a_1 = \sigma(z_1)$ 2,10

$z_2 = w_2 a_1 + b_2, \quad \hat{y} = \sigma(z_2) = 1/1+e^{-z_2}$

$$\frac{\partial J}{\partial z_2} = \frac{\partial J}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z_2} = \left(\frac{-y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\right) \cdot \frac{e^{-z_2}}{(1+e^{-z_2})^2} =$$

$$\left(\frac{-y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\right) \cdot \hat{y}(1-\hat{y})$$

$$\frac{\partial z_2}{\partial a_1} = \frac{\partial\, w_2 a_1 + b_2}{\partial a_1} = w_2$$

$$\frac{\partial a_1}{\partial z_1} = \frac{\partial\, \mathrm{Relu}(z_1)}{\partial z_1} = \begin{cases} 1 & z_1 > 0 \\ 0 & z_1 \leq 0 \end{cases} = \mathrm{Relu}'(z_1)$$

$$\frac{\partial z_1}{\partial w_1} = \frac{\partial\, w_1 x + b_1}{\partial w_1} = x$$

$$\frac{\partial J}{\partial W_1} = \left(\frac{-y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\right)\cdot \hat{y}(1-\hat{y})\cdot W_2 \cdot \mathrm{Relu}'(z_1)\cdot X$$

$$= (\hat{y}-y)\, W_2 \cdot \mathrm{Relu}'(z_1)\cdot X$$

$$\frac{\partial J}{\partial b_1} = \frac{\partial J}{\partial z_2}\cdot\frac{\partial z_2}{\partial a_1}\cdot\frac{\partial a_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial b_1}$$

שרשרת

$$\frac{\partial J}{\partial z_2} = \left(\frac{-y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\right)\cdot \hat{y}\cdot(1-\hat{y}) = \hat{y}-y$$

$$\frac{\partial z_2}{\partial a_1} = \frac{W_2\cdot a_1 + b_2 + X}{\partial a_1} = W_2$$

$$\frac{\partial a_1}{\partial z_1} = \frac{\partial\, \mathrm{Relu}(z_1)}{\partial z_1} = \mathrm{Relu}'(z_1)$$

$$\frac{\partial z_1}{\partial b_1} = \frac{\partial\, W_1 X + b_1}{\partial b_1} = 1$$

לא נדרש

$$\frac{\partial J}{\partial b_1} = \left(\frac{-y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\right)\hat{y}(1-\hat{y})\cdot W_2\cdot \mathrm{Relu}'(z_1) =$$

$$(\hat{y}-y)\, W_2\cdot \mathrm{Relu}'(z_1)$$

$$\frac{\partial J}{\partial X} = \frac{\partial J}{\partial z_2}\cdot\frac{\partial z_2}{\partial a_1}\cdot\frac{\partial a_1}{\partial z_1}\cdot\frac{\partial z_1}{\partial X}$$

$$\frac{\partial}{\partial x} = \frac{\partial z_2}{\partial a_1} \cdot \frac{\partial}{\partial z_1} \cdot \frac{\partial}{\partial x}$$

$$\frac{\partial z_1}{\partial x} = \frac{\partial\, w_1 x + b_1}{\partial x} = w_1$$

$$\frac{\partial J}{\partial x} = \left(\frac{-y}{\hat{y}} + \frac{1-y}{1-\hat{y}}\right)\cdot \hat{y}\cdot(1-\hat{y})\cdot w_2 \cdot \mathrm{Relu}'(z_1)\cdot w_1$$

$$= (\hat{y}-y)\, w_2\, \mathrm{Relu}'(z_1)\, w_1$$

B.4. אלו מושגים דיי"ר בכיתה לשמור עליו

מושגים אלו?

הינו בכיתה לשמור את המשתנים עליו כל המושג:

$\hat{y}: z_2$, $\mathrm{Relu}(z_1)': z_1$, $\frac{\partial J}{\partial w_2}: \hat{y}$, $\mathrm{Relu}(z_1): a_1$,

$\frac{\partial J}{\partial b_2}: \hat{y}$, $\frac{\partial J}{\partial w_1}: \hat{y}$, $\mathrm{Relu}'(z_1), x$, $\frac{\partial J}{\partial b_1}: \hat{y}, \mathrm{Relu}'(z_1)$,

$\frac{\partial J}{\partial x}: \hat{y}, \mathrm{Relu}'(z_1)$

לסכום: $a_1, z_1, z_2, \hat{y}, \mathrm{Relu}(z_1)$

דוגמה, שמירת הנגזרות המלאות בכל קשר

כפי שאנו רואים הקודמים: $\frac{\partial J}{\partial x}, \frac{\partial J}{\partial b_2}, \frac{\partial J}{\partial w_2}, \frac{\partial J}{\partial b_1}, \frac{\partial J}{\partial w_1}$

$$\cancel{\frac{\partial \hat{y}}{\partial x} = \frac{\partial\left(\overbrace{w_2 \cdot \mathrm{Relu}\left(\overbrace{w_1 x + b_1}^{z_1}\right)}^{a_1} + \overbrace{x + b_2}^{z_2}\right)}{\partial x}} =$$

$$\cancel{\frac{\partial \hat{y}}{\partial x} = \frac{\partial \sigma(w_2 \cdot \mathrm{Relu}(w_1 x + b_1) + x \cdot b_2)}{\partial x} =}$$

$$\cancel{\frac{\partial J}{\partial x} = \sigma'(z_2)\Big[\underbrace{w_2 \cdot \mathrm{Relu}'(w_1 x + b_1) \cdot w_1 + 1}_{\frac{\partial z_2}{\partial x}}\Big] =}$$

$$\cancel{\boxed{\sigma'(z_2) \cdot w_2 \cdot \mathrm{Relu}'(z_1) \cdot w_1 + 1} = \frac{\partial \hat{y}}{\partial x}}$$

$$\cancel{J = -y \log(\hat{y}) - (1-y)\log(1-\hat{y})}$$

$$\therefore \frac{\partial J}{\partial x} = \underbrace{\frac{\partial J}{\partial z_2}}_{\hat{y}-y} \cdot \underbrace{\frac{\partial z_2}{\partial x}}_{\left[1 + \frac{\partial z_1}{\partial x}\right] = \left[1 + w_1 \cdot w_2 \cdot (\hat{y}-y) \cdot \mathrm{Relu}'(z_1)\right]} = (\hat{y}-y)\left[1 + w_1 w_2 (\hat{y}-y)\,\mathrm{Relu}'(z_1)\right]$$