Matrix Calculus Practice Questions

2023-07-09

Questions compiled by Salman Naqvi. All questions in this paper are based of the The Matrix Calculus You Need For Deep Learning paper by Terence Parr and Jeremy Howard. Do let me know of any corrections.

Questions 1
Answers 5

Questions

1. Find $\frac{\mathrm{d}}{\mathrm{d}x}9(x+x^2)$.

2. $f(x,y)=3x^2y$. Find $\nabla f(x,y)$.

3. $f(x,y)=3x^2y$ and $g(x,y)=2x+y^8$.

    a. Find the Jacobian containing $\nabla f(x,y)$ and $\nabla g(x,y)$ in numerator layout.

    b. Find the Jacobian containing $\nabla f(x,y)$ and $\nabla g(x,y)$ in denominator layout.

4. Let $\vec{\mathrm{y}}=\vec{\mathrm{f}}(\vec{\mathrm{x}})=f_i(x_i)=x_i$. Show that $\frac{\partial\vec{\mathrm{y}}}{\partial\vec{\mathrm{x}}}=I$.

5. $\vec{\mathrm{y}}=\vec{\mathrm{f}}(\vec{\mathrm{w}})\bigcirc\vec{\mathrm{g}}(\vec{\mathrm{x}})=f_i(w_i)\bigcirc g_i(x_i)$. Prove the following identities.

    a. If $\vec{\mathrm{y}}=\vec{\mathrm{f}}(\vec{\mathrm{w}})\oplus\vec{\mathrm{g}}(\vec{\mathrm{w}})$,

        i. show that $\frac{\partial(\vec{\mathrm{w}}\oplus\vec{\mathrm{x}})}{\partial\vec{\mathrm{w}}}=I$.

        ii. show that $\frac{\partial(\vec{\mathrm{w}}\oplus\vec{\mathrm{x}})}{\partial\vec{\mathrm{x}}}=I$.

    b. If $\vec{\mathrm{y}}=\vec{\mathrm{f}}(\vec{\mathrm{w}})\ominus\vec{\mathrm{g}}(\vec{\mathrm{w}})$,

        i. show that $\frac{\partial(\vec{\mathrm{w}}\ominus\vec{\mathrm{x}})}{\partial\vec{\mathrm{w}}}=I$.

        ii. show that $\frac{\partial(\vec{\mathrm{w}}\ominus\vec{\mathrm{x}})}{\partial\vec{\mathrm{x}}}=-I$.

    c. If $\vec{\mathrm{y}}=\vec{\mathrm{f}}(\vec{\mathrm{w}})\otimes\vec{\mathrm{g}}(\vec{\mathrm{w}})$,

        i. show that $\frac{\partial(\vec{\mathrm{w}}\otimes\vec{\mathrm{x}})}{\partial\vec{\mathrm{w}}}=\mathrm{diag}(\vec{\mathrm{x}})$.

ii. show that $\frac{\partial(\vec{\mathrm{w}}\otimes\vec{\mathrm{x}})}{\partial\vec{\mathrm{x}}} = \mathrm{diag}(\vec{\mathrm{w}})$.

d. If $\vec{\mathrm{y}} = \vec{\mathrm{f}}(\vec{\mathrm{w}}) \oslash \vec{\mathrm{g}}(\vec{\mathrm{w}})$,

i. show that $\frac{\partial(\vec{\mathrm{w}}\oslash\vec{\mathrm{x}})}{\partial\vec{\mathrm{w}}} = \mathrm{diag}\left(\cdots, \frac{1}{x_i}, \cdots\right)$.

ii. show that $\frac{\partial(\vec{\mathrm{w}}\oslash\vec{\mathrm{x}})}{\partial\vec{\mathrm{x}}} = \mathrm{diag}(\cdots, -\frac{w_i}{x_i^2}$.

6. $\vec{\mathrm{y}} = \vec{\mathrm{f}}(\vec{\mathrm{x}}) \bigcirc \vec{\mathrm{g}}(z) = \vec{\mathrm{x}} \bigcirc \vec{1}z$. Prove the following identities once with matricies and once with a general form of the equation.

a. $\frac{\partial(\vec{\mathrm{x}}\oplus z)}{\partial\vec{\mathrm{x}}} = I$

b. $\frac{\partial(\vec{\mathrm{x}}\oplus z)}{\partial z} = \vec{1}$

c. $\frac{\partial(\vec{\mathrm{x}}\otimes z)}{\partial\vec{\mathrm{x}}} = Iz$

d. $\frac{\partial(\vec{\mathrm{x}}\otimes z)}{\partial z} = \vec{\mathrm{x}}$

7. $y = \mathrm{sum}\left(\vec{\mathrm{f}}(\vec{\mathrm{x}})\right) = \sum_{i=1}^{n} f_i(\vec{\mathrm{x}})$, where $f_i(\vec{\mathrm{x}}) \neq x_i$. Show that $\frac{\partial y}{\partial\vec{\mathrm{x}}} = \left[\sum_i \frac{\partial f_{i(\vec{\mathrm{x}})}}{\partial x_1} \ \sum_i \frac{\partial f_{i(\vec{\mathrm{x}})}}{\partial x_2} \ \cdots \ \sum_i \frac{\partial f_{i(\vec{\mathrm{x}})}}{\partial x_n}\right]$.

8. $y = \mathrm{sum}(\vec{\mathrm{x}})$. Show that $\nabla y = \vec{1}^T$.

9. $y = \mathrm{sum}(\vec{\mathrm{x}}z)$.

a. Show that $\frac{\partial y}{\partial\vec{\mathrm{x}}} = \vec{1}^T z$.

b. Show that $\frac{\partial y}{\partial z} = \mathrm{sum}(\vec{\mathrm{x}})$.

10. Find $\frac{\mathrm{d}}{\mathrm{d}x}\sin(x^2)$.

11. $y = \ln(\sin^2(x^3))$. Find $\frac{\mathrm{d}y}{\mathrm{d}x}$.

12. $y = f(x) = x + x^2$. Solve for $\frac{\mathrm{d}y}{\mathrm{d}x}$ using total derivatives.

13. Show that $\frac{\partial f(x_1, u_1, \ldots, u_n)}{\partial x} = \frac{\partial f}{\partial x} + \sum_{i=1}^{n} \frac{\partial f}{\partial u_i}\frac{\partial u_i}{\partial x}$, where $u_1, u_2, \ldots, u_n$ are all functions of x.

a. Further simplify $\frac{\partial f}{\partial x} + \sum_{i=1}^{n} \frac{\partial f}{\partial u_i}\frac{\partial u_i}{\partial x}$ to $\sum_{i=1}^{n+1} \frac{\partial f}{\partial u_i}\frac{\partial u_i}{\partial x}$.

14. $f(x) = \sin(x + x^2)$. Solve for $\frac{\partial f(x)}{\partial x}$.

15. $y = x \cdot x^2$. Solve for $\frac{\mathrm{d}y}{\mathrm{d}x}$ using total derivatives.

16. $\vec{y} = \vec{f}(x) = \begin{bmatrix} y_1(x) \\ y_2(x) \end{bmatrix} = \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} = \begin{bmatrix} \ln(x^2) \\ \sin(3x) \end{bmatrix}$. Find $\frac{\partial \vec{y}}{\partial x}$.

17. $\vec{y} = \vec{f}(\vec{g}(x))$. Show that $\frac{\partial \vec{y}}{\partial x} = \frac{\partial \vec{f}}{\partial \vec{g}} \frac{\partial \vec{g}}{\partial x}$.

18. $\vec{y} = \vec{f}(\vec{g}(\vec{x}))$, where $f_i(\vec{x}) = g_i(\vec{x}) = x_i$.

    a. Show that $\frac{\partial \vec{f}}{\partial \vec{g}} = \operatorname{diag}\left(\frac{\partial f_i}{\partial g_i}\right)$.

    b. Show that $\frac{\partial \vec{g}}{\partial \vec{x}} = \operatorname{diag}\left(\frac{\partial g_i}{\partial x_i}\right)$.

    c. Show that $\frac{\partial}{\partial \vec{x}} \vec{f}(\vec{g}(\vec{x})) = \operatorname{diag}\left(\frac{\partial f_i}{\partial g_i} \frac{\partial g_i}{\partial x_i}\right)$

19. $y = \max(0, \vec{w} \cdot \vec{x} + b)$.

    a. Find $\frac{\partial(\vec{w} \otimes \vec{x})}{\partial \vec{w}}$.

    b. Now find $\frac{\partial \operatorname{sum}(\vec{u})}{\partial \vec{u}}$, where $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$.

    c. Therefore, find

        i. $\frac{\partial(\vec{w} \cdot \vec{x} + b)}{\partial \vec{w}}$.

        ii. $\frac{\partial(\vec{w} \cdot \vec{x} + b)}{\partial b}$.

    d. Let $z = \vec{w} \cdot \vec{x} + b, \therefore y = \max(0, z)$. Find $\frac{\partial y}{\partial z}$.

    e. Finally, find

        i. $\frac{\partial y}{\partial \vec{w}}$.

        ii. $\frac{\partial y}{\partial b}$.

20. The MSE (Mean Squared Error) for two values is given by $\frac{(\hat{y} - y)^2}{2}$, where y denotes a prediction and $\hat{y}$ denotes the corresponding target.

    a. If we have multiple data samples that are stored in another vector, $\vec{X} = \begin{bmatrix} \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_n \end{bmatrix}^T$, and the targets for each sample are stored in $\vec{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^T$, and that a prediction is given by $\max(\vec{w} \cdot \vec{x}_i + b)$, show that the MSE simplifies to $\frac{1}{N} \sum_{i=1}^{N} (y_i - \max(0, \vec{w} \cdot \vec{x}_i + b))^2$, where $N = |X|$.

b. Find $\frac{\partial(\vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_i + b)}{\partial \vec{\mathrm{w}}}$.

c. Find $\frac{\partial \max(\vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_i + b)}{\partial \vec{\mathrm{w}}}$.

d. Find $\frac{\partial(y_i - \max(\vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_i + b))}{\partial \vec{\mathrm{w}}}$.

e. Find $\frac{\partial(y_i - \max(\vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_i + b))^2}{\partial \vec{\mathrm{w}}}$.

f. Finally, find $\frac{\partial \text{ MSE}}{\partial \vec{\mathrm{w}}} = \frac{\partial}{\partial \vec{\mathrm{w}}} \left(\frac{1}{N} \sum_{i=1}^{N} (y_i - \max(0, \vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_i + b))^2 \right)$. Simplify the answer by letting $e_i = \vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_i + b - y_i$.

g. Similarly, solve for $\frac{\partial \text{ MSE}}{\partial b}$ and simplify with e_i.

h. Instead of finding the partial derivative with respect to $\vec{\mathrm{w}}$ and b separately, we can let $\hat{\mathrm{w}} = \begin{bmatrix} \vec{\mathrm{w}}^T & b \end{bmatrix}^T$ and let $\hat{\mathrm{x}}_i = \begin{bmatrix} \vec{\mathrm{x}}_i^T & 1 \end{bmatrix}$, and instead solve for $\frac{\partial \text{ MSE}}{\partial \hat{\mathrm{w}}}$. Solve for $\frac{\partial \text{ MSE}}{\partial \hat{\mathrm{w}}}$ and simplify by letting $e_i = \hat{\mathrm{w}} \cdot \hat{\mathrm{x}}_i - y$.

Answers

1. $\frac{\mathrm{d}}{\mathrm{d}x} 9(x + x^2) = 9 + 18x$

2. $\nabla f(x, y) = \begin{bmatrix} 6xy \\ 3x^2 \end{bmatrix}$

3.

   a. $J = \begin{bmatrix} 6xy & 3x^2 \\ 2 & 8y^7 \end{bmatrix}$

   b. $J = \begin{bmatrix} 6xy & 2 \\ 3x^2 & 8y^7 \end{bmatrix}$

10. $\frac{\mathrm{d}}{\mathrm{d}x} \sin(x^2) = 2x \cos(x^2)$

11. $\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{6x^2 \cos(x^3)}{\sin(x^3)}$

12. $\frac{\mathrm{d}y}{\mathrm{d}x} = 1 + 2x$

13.

    a. Let $x = u_{n+1}$. Then $\frac{\partial f(u_1, \ldots, u_{n+1})}{\partial x} = \sum_{i=1}^{n+1} \frac{\partial f}{\partial u_i} \frac{\partial u_i}{\partial x}$.

14. $\frac{\mathrm{d}f(x)}{\mathrm{d}x} = \cos(x + x^2)(1 + 2x)$

15. $\frac{\mathrm{d}y}{\mathrm{d}x} = 3x^2$

17. $\frac{\mathrm{d}\vec{\mathrm{y}}}{\mathrm{d}x} = \begin{bmatrix} \frac{2}{x} \\ 3\cos(3x) \end{bmatrix}$

19.

    a. $\frac{\partial(\vec{\mathrm{w}} \otimes \vec{\mathrm{x}})}{\partial \vec{\mathrm{w}}} = \mathrm{diag}(\vec{\mathrm{x}})$

b. $\frac{\partial \text{ sum}(\vec{\mathrm{u}})}{\partial \vec{\mathrm{u}}} = \vec{1}^T$

c. i. $\frac{\partial y}{\partial \vec{\mathrm{w}}} = \vec{\mathrm{x}}^T$

ii. $\frac{\partial y}{\partial b} = 1$

d. $\frac{\partial y}{\partial z} = \begin{cases} 0 & z \leq 0 \\ 1 & z > 0 \end{cases}$

e. i. $\frac{\partial y}{\partial \vec{\mathrm{w}}} = \begin{cases} \vec{0}^T & \vec{\mathrm{w}} \cdot \vec{\mathrm{x}} + b \leq 0 \\ \vec{\mathrm{x}}^T & \vec{\mathrm{w}} \cdot \vec{\mathrm{x}} + b > 0 \end{cases}$

ii. $\frac{\partial y}{\partial b} = \begin{cases} 0 & \vec{\mathrm{w}} \cdot \vec{\mathrm{x}} + b \leq 0 \\ 1 & \vec{\mathrm{w}} \cdot \vec{\mathrm{x}} + b > 0 \end{cases}$

20.

a. For a single sample, MSE $= \frac{(y - (\vec{\mathrm{w}} \cdot \vec{\mathrm{x}} + b))^2}{2}$. Therefore, for multiple samples,

MSE $= \frac{1}{N} \sum_{i=1}^{N} (y_i - (\vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_i + b))^2$.

b. $\frac{\partial (\vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_i + b)}{\partial \vec{\mathrm{w}}} = \vec{\mathrm{x}}_i^T$

c. $\frac{\partial \max(0, \vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_i + b)}{\partial \vec{\mathrm{w}}} = \begin{cases} \vec{0}^T & \vec{\mathrm{w}} \cdot \vec{\mathrm{x}} + b \\ \vec{\mathrm{x}}_i^T & \vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_{\mathrm{i}} + b \end{cases}$

d. $\frac{\partial (y_i - \max(0, \vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_{\mathrm{i}} + b))}{\partial \vec{\mathrm{w}}} = \begin{cases} \vec{0}^T & \vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_{\mathrm{i}} + b \leq 0 \\ -\vec{\mathrm{x}}_i^T & \vec{\mathrm{w}} \cdot \vec{\mathrm{x}} + b > 0 \end{cases}$

e. $\frac{\partial (y_i - \max(0, \vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_{\mathrm{i}} + b))^2}{\partial \vec{\mathrm{w}}} = \begin{cases} \vec{0}^T & \vec{\mathrm{w}} \cdot \vec{\mathrm{x}} + b \leq 0 \\ -2(y_i - \max(\vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_{\mathrm{i}} + b))\vec{\mathrm{x}}_i^T & \vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_{\mathrm{i}} + b > 0 \end{cases}$

f. $\frac{\partial}{\partial \vec{\mathrm{w}}} \left(\frac{1}{N} \sum_{i=1}^{N} (y_i - \max(0, \vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_{\mathrm{i}} + b))^2 \right) = \begin{cases} \vec{0}^T & \vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_{\mathrm{i}} + b \leq 0 \\ \frac{2}{N} \sum_{i=1}^{N} e_i \vec{\mathrm{x}}_i^T & \vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_{\mathrm{i}} + b > 0 \end{cases}$

g. $\frac{\partial}{\partial b} \left(\frac{1}{N} \sum_{i=1}^{N} (y_i - \max(0, \vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_{\mathrm{i}} + b))^2 \right) = \begin{cases} \vec{0}^T & \vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_{\mathrm{i}} + b \leq 0 \\ \frac{2}{N} \sum_{i=1}^{N} e_i & \vec{\mathrm{w}} \cdot \vec{\mathrm{x}}_{\mathrm{i}} + b > 0 \end{cases}$

h. $\frac{\partial}{\partial \hat{\mathrm{w}}} \left(\frac{1}{N} \sum_{i=1}^{N} (y_i - \max(0, \hat{\mathrm{w}} \cdot \hat{\mathrm{x}}))^2 \right) = \begin{cases} \vec{0}^T & \hat{\mathrm{w}} \cdot \hat{\mathrm{x}} \leq 0 \\ \frac{2}{N} \sum_{i=1}^{N} e_i \hat{\mathrm{x}}_i^T & \hat{\mathrm{w}} \cdot \hat{\mathrm{x}} > 0 \end{cases}$