CS 316: Introduction to Deep Learning

Matrix Calculus and Chain Rule
Week 3

Dr Abdul Samad

Lecture Outline

- Scalar derivatives
- Vector calculus and partial derivatives
- Chain rule
- Matrix calculus

Common Derivatives

- **Derivative of constants.** $\frac{d}{dx}c = 0$.
- **Derivative of linear functions.** $\frac{d}{dx}(ax) = a$.
- **Power rule.** $\frac{d}{dx}x^n = nx^{n-1}$.
- **Derivative of exponentials.** $\frac{d}{dx}e^x = e^x$.
- **Derivative of the logarithm.** $\frac{d}{dx}\log(x) = \frac{1}{x}$.

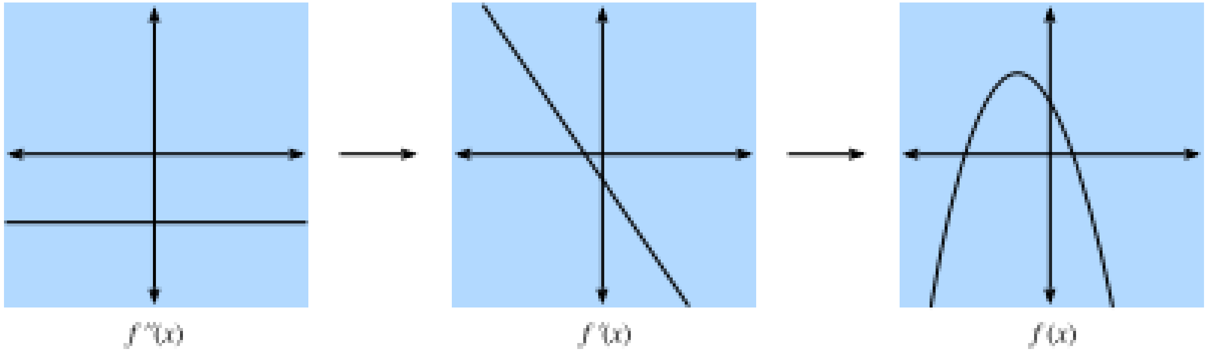
Derivative Rules

- **Sum rule.** $\frac{d}{dx}\left(g(x)+h(x)\right)=\frac{dg}{dx}(x)+\frac{dh}{dx}(x)$.
- **Product rule.** $\frac{d}{dx}\left(g(x)\cdot h(x)\right)=g(x)\frac{dh}{dx}(x)+\frac{dg}{dx}(x)h(x)$.
- **Chain rule.** $\frac{d}{dx}g(h(x))=\frac{dg}{dh}(h(x))\cdot\frac{dh}{dx}(x)$.

Higher order derivatives

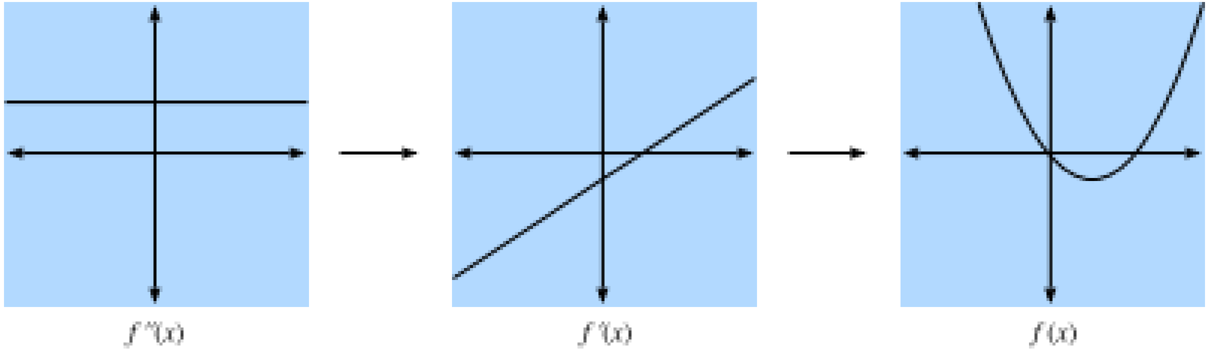
$$f^{(n)}(x) = \frac{d^n f}{dx^n} = \left(\frac{d}{dx}\right)^n f.$$

Visualising Derivatives - 1



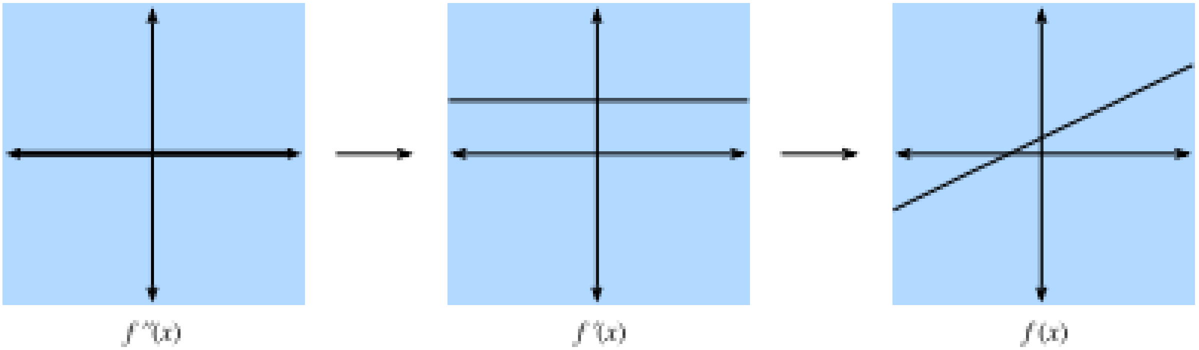
The second derivative is a negative constant, then the first derivative is decreasing, which implies the function itself has a maximum.

Visualising Derivatives - 2



The second derivative is a positive constant, then the first derivative is increasing, which implies the function itself has a minimum.

Visualising Derivatives - 3



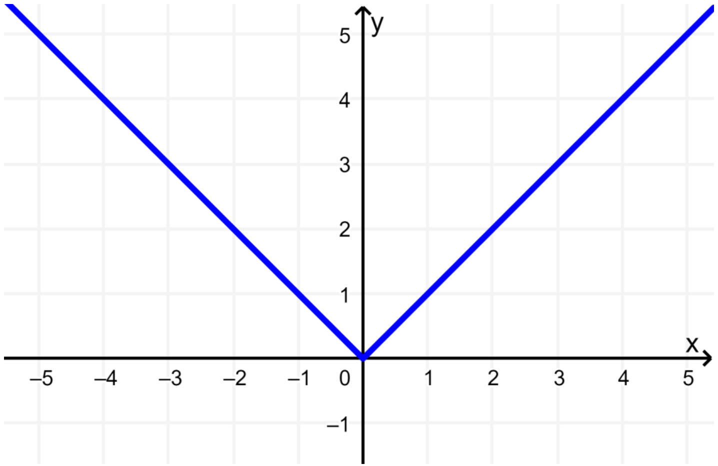
The second derivative is zero, then the first derivative is constant, which implies the function itself is a straight line.

Sub derivative

- Extend derivative to non-differentiable cases

Example - 1

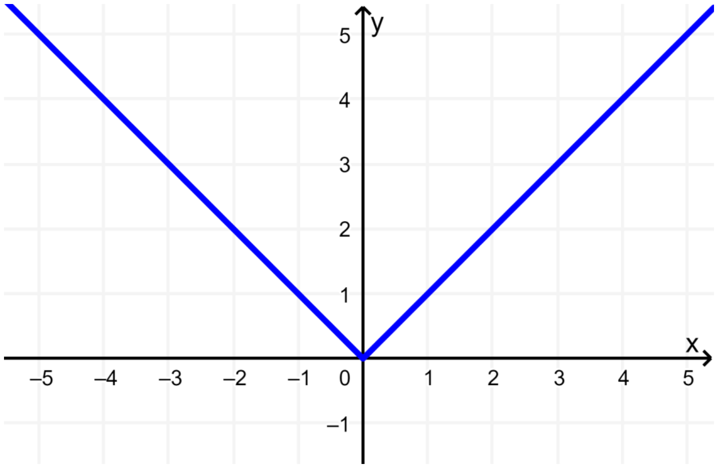
Compute the derivative of $f(x) = |x|$ w.r.t x.



Example - 1

Compute the derivative of $f(x) = |x|$ w.r.t x.

slope= - 1

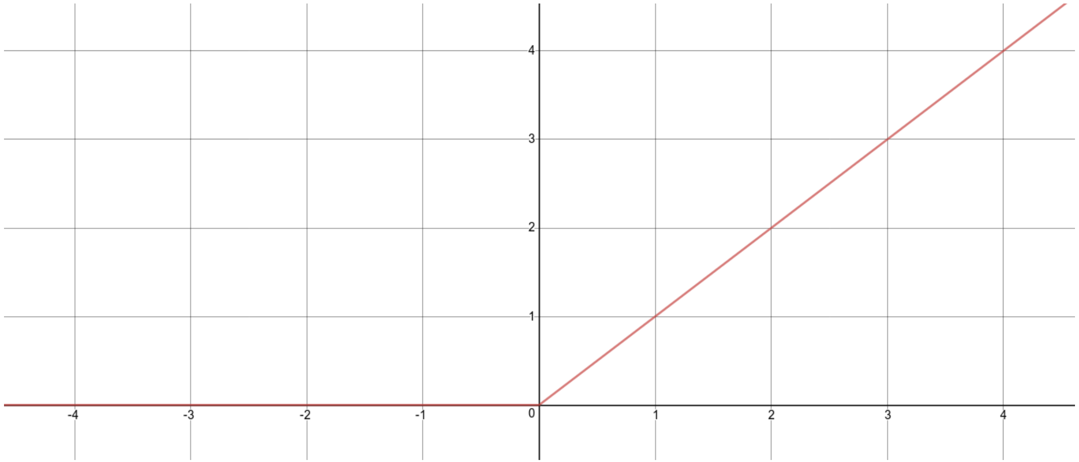


slope= 1

$$\frac{\partial |x|}{\partial x} = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \\ a & \text{if } x = 0, \quad a \in [-1,1] \end{cases}$$

Example - 2

Compute the derivative of $f(x) = max(0, x)$ w.r.t x.



Example - 2

Compute the derivative of $f(x) = max(0, x)$ w.r.t x.

slope= 0

slope= 1

$$\frac{\partial}{\partial x}\max(x,0) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x < 0 \\ a & \text{if } x = 0, \quad a \in [0,1] \end{cases}$$

Vector Calculus

- Generalize derivatives into vectors.

		Scalar	Vector
		x	$\mathbf{x}$
Scalar	y	$\frac{\partial y}{\partial x}$	$\frac{\partial y}{\partial \mathbf{x}}$
Vector	$\mathbf{y}$	$\frac{\partial \mathbf{y}}{\partial x}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$

Example - 3

Compute the derivative of $f(x_1, x_2) = x_1^2 + 2x_2^2$

Example - 3

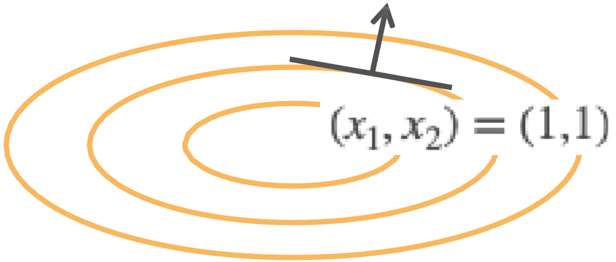
Compute the derivative of $f(x_1, x_2) = x_1^2 + 2x_2^2$

$$\frac{\delta f}{\delta x_1} = 2x_1$$

$$\frac{\delta f}{\delta x_2} = 4x_2$$

$$\nabla f(x_1, x_2) = [2x_1, 4x_2]$$

Direction (2, 4), perpendicular to the contour lines



$\partial y / \partial \mathbf{x}$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \frac{\partial y}{\partial \mathbf{x}} = \left[\frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \ldots, \frac{\partial y}{\partial x_n} \right]$$

	x	$\mathbf{x}$
y	$\frac{\partial y}{\partial x}$	$\frac{\partial y}{\partial \mathbf{x}}$
$\mathbf{y}$	$\frac{\partial \mathbf{y}}{\partial x}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$

X is a vector and y is a scalar.

Common Derivatives

- Derivative of constant $\frac{\delta}{\delta \mathbf{x}} a = \mathbf{0}^T$
- Derivative of linear function $\frac{\delta}{\delta \mathbf{x}} a\mathbf{u} = a\frac{\delta \mathbf{u}}{\delta \mathbf{x}}$
- Derivative of sum $\frac{\delta}{\delta \mathbf{x}} sum(\mathbf{x}) = \mathbf{1}^T$
- Derivative of L2 norm $\frac{\delta}{\delta \mathbf{x}} ||\mathbf{x}||_2^2 = 2\mathbf{x}^T$

Derivative Rules

- Sum rule $$\frac{\delta}{\delta \mathbf{x}}(\mathbf{u}+\mathbf{v}) = \frac{\delta \mathbf{u}}{\delta \mathbf{x}} + \frac{\delta \mathbf{v}}{\delta \mathbf{x}}$$
- Product rule $$\frac{\delta}{\delta \mathbf{x}}(\mathbf{u}\mathbf{v}) = \frac{\delta \mathbf{u}}{\delta \mathbf{x}}\mathbf{v} + \frac{\delta \mathbf{v}}{\delta \mathbf{x}}\mathbf{u}$$
- Dot product $$\frac{\delta}{\delta \mathbf{x}}\langle \mathbf{u}, \mathbf{v}\rangle = \mathbf{v}^T \frac{\delta \mathbf{u}}{\delta \mathbf{x}} + \mathbf{u}^T \frac{\delta \mathbf{v}}{\delta \mathbf{x}}$$

$\partial \mathbf{y} / \partial x$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$$\frac{\partial \mathbf{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} \\ \frac{\partial y_2}{\partial x} \\ \vdots \\ \frac{\partial y_m}{\partial x} \end{bmatrix}$$

	x	$\mathbf{x}$
y	$\frac{\partial y}{\partial x}$	$\frac{\partial y}{\partial \mathbf{x}}$
$\mathbf{y}$	$\frac{\partial \mathbf{y}}{\partial x}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$

X is a scalar and y is a vector.

$\partial \mathbf{y} / \partial \mathbf{x}$

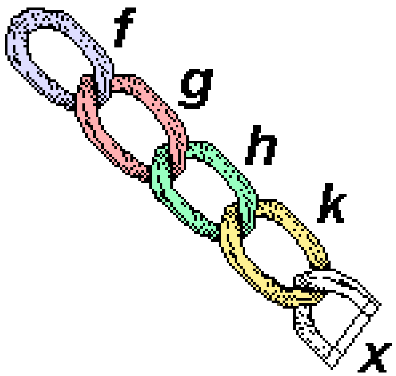
	x	$\mathbf{x}$
y	$\frac{\partial y}{\partial x}$	$\frac{\partial y}{\partial \mathbf{x}}$
$\mathbf{y}$	$\frac{\partial \mathbf{y}}{\partial x}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial \mathbf{x}} \\ \frac{\partial y_2}{\partial \mathbf{x}} \\ \vdots \\ \frac{\partial y_m}{\partial \mathbf{x}} \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1}, \frac{\partial y_1}{\partial x_2}, \ldots, \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1}, \frac{\partial y_2}{\partial x_2}, \ldots, \frac{\partial y_2}{\partial x_n} \\ \vdots \\ \frac{\partial y_m}{\partial x_1}, \frac{\partial y_m}{\partial x_2}, \ldots, \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

Both x and y are vectors.

Chain Rule



Generalize to vectors

- Chain rule for scalars:

$$y = f(u),\ u = g(x) \qquad \frac{\partial y}{\partial x} = \frac{\partial y}{\partial u}\frac{\partial u}{\partial x}$$

- Generalize to vectors straightforwardly

$$\underset{(1,n)}{\frac{\partial y}{\partial \mathbf{x}}} = \underset{(1,)}{\frac{\partial y}{\partial u}}\underset{(1,n)}{\frac{\partial u}{\partial \mathbf{x}}} \qquad \underset{(1,n)}{\frac{\partial y}{\partial \mathbf{x}}} = \underset{(1,k)}{\frac{\partial y}{\partial \mathbf{u}}}\underset{(k,n)}{\frac{\partial \mathbf{u}}{\partial \mathbf{x}}} \qquad \underset{(m,n)}{\frac{\partial \mathbf{y}}{\partial \mathbf{x}}} = \underset{(m,k)}{\frac{\partial \mathbf{y}}{\partial \mathbf{u}}}\underset{(k,n)}{\frac{\partial \mathbf{u}}{\partial \mathbf{x}}}$$

Matrix Calculus

- Derivative of constant $\frac{\delta}{\delta \mathbf{x}} a = \mathbf{0}$
- Derivative of identity function $\frac{\delta}{\delta \mathbf{x}} \mathbf{x} = \mathbf{I}$
- Derivative of linear function $\frac{\delta}{\delta \mathbf{x}} \mathbf{A}\mathbf{x} = \mathbf{A}$
- Derivative of transpose $\frac{\delta}{\delta \mathbf{x}} \mathbf{x}^T \mathbf{A} = \mathbf{A}^T$

Generalize to Matrices

		Scalar	Vector	Matrix
		x $(1,)$	$\mathbf{x}$ $(n,1)$	$\mathbf{X}$ (n,k)
Scalar	y $(1,)$	$\frac{\partial y}{\partial x}$ $(1,)$	$\frac{\partial y}{\partial \mathbf{x}}$ $(1,n)$	$\frac{\partial y}{\partial \mathbf{X}}$ (k,n)
Vector	$\mathbf{y}$ $(m,1)$	$\frac{\partial \mathbf{y}}{\partial x}$ $(m,1)$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ (m,n)	$\frac{\partial \mathbf{y}}{\partial \mathbf{X}}$ (m,k,n)
Matrix	$\mathbf{Y}$ (m,l)	$\frac{\partial \mathbf{Y}}{\partial x}$ (m,l)	$\frac{\partial \mathbf{Y}}{\partial \mathbf{x}}$ (m,l,n)	$\frac{\partial \mathbf{Y}}{\partial \mathbf{X}}$ (m,l,k,n)

Example - 4

Compute $\frac{\delta z}{\delta w}$ where $z = (\langle x, w \rangle - y)^2$. Assume $x, w \in \mathbb{R}^n, y \in \mathbb{R}$

Example - 4

Compute $\frac{\delta z}{\delta w}$ where $z = (\langle x, w \rangle - y)^2$. Assume $x, w \in \mathbb{R}^n, y \in \mathbb{R}$

$$a = \langle x, w \rangle$$

$$b = a - y$$

$$z = b^2$$

$$\frac{\delta z}{\delta w} = \frac{\delta z}{\delta b} \times \frac{\delta b}{\delta a} \times \frac{\delta a}{\delta w}$$

$$\frac{\delta z}{\delta b} = 2b$$

$$\frac{\delta b}{\delta a} = 1$$

$$\frac{\delta a}{\delta w} = x^T$$

$$\frac{\delta z}{\delta w} = 2b \cdot 1 \cdot x^T$$

$$\frac{\delta z}{\delta w} = 2(\langle x, w \rangle - y) \cdot x^T$$

Example - 5

Compute $\dfrac{\delta z}{\delta w}$ where $z = \|Xw - y\|^2$. Assume $X \in \mathbb{R}^{m \times n}, w \in \mathbb{R}^n, y \in \mathbb{R}^m$

Example - 5

Compute $\frac{\delta z}{\delta w}$ where $z = \|Xw - y\|^2$. Assume $X \in \mathbb{R}^{m \times n}, w \in \mathbb{R}^n, y \in \mathbb{R}^m$

$$a = Xw$$

$$b = a - y$$

$$z = \|b\|^2$$

$$\frac{\delta z}{\delta w} = \frac{\delta z}{\delta b} \times \frac{\delta b}{\delta a} \times \frac{\delta a}{\delta w}$$

$$\frac{\delta z}{\delta b} = 2b^T$$

$$\frac{\delta b}{\delta a} = I$$

$$\frac{\delta a}{\delta w} = X$$

$$\frac{\delta z}{\delta w} = 2b^T \cdot I \cdot X$$

$$\frac{\delta z}{\delta w} = 2(Xw - y)^T \cdot X$$