Machine Learning

Calculus - A crash course

Kia Nazarpour

Based on Hao Tang's slides

Learning Outcomes

1. Remember key concepts about derivatives
2. Feel confident about a source of needed mathematics to review
3. Warm up with some Calculus examples

Derivative in 1D

Derivative is the results of differentiation.
The derivative of a function $f : \mathbb{R} \to \mathbb{R}$ at x_0 is

$$\left(\frac{d}{dx} f\right)(x_0) \quad = \quad \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

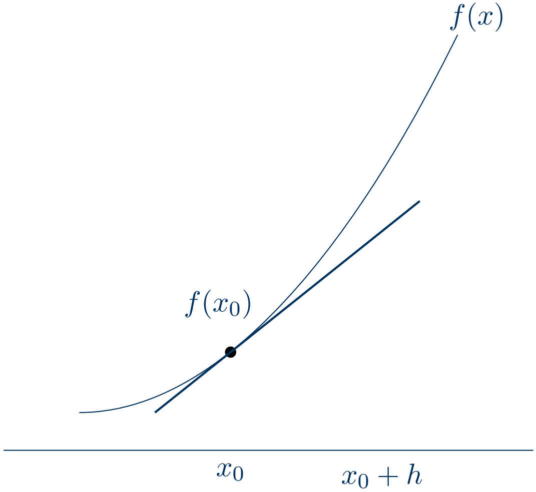
Derivative in 1D

Derivative is the results of differentiation.
The derivative of a function $f : \mathbb{R} \to \mathbb{R}$ at x_0 is

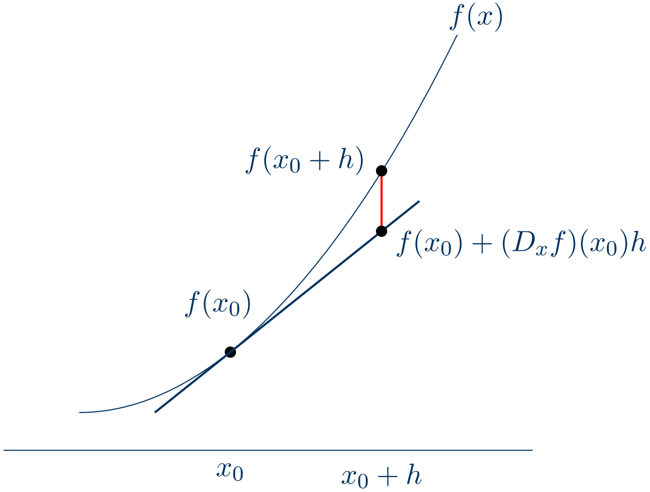
$$\begin{aligned}\left(\frac{d}{dx}f\right)(x_0) &= \lim_{h\to 0}\frac{f(x_0+h)-f(x_0)}{h} \\ &= (D_x f)(x_0)\end{aligned}$$

Conventionally, we understand this equation as a rate of change at x_0. We would like to offer a different perspective here.

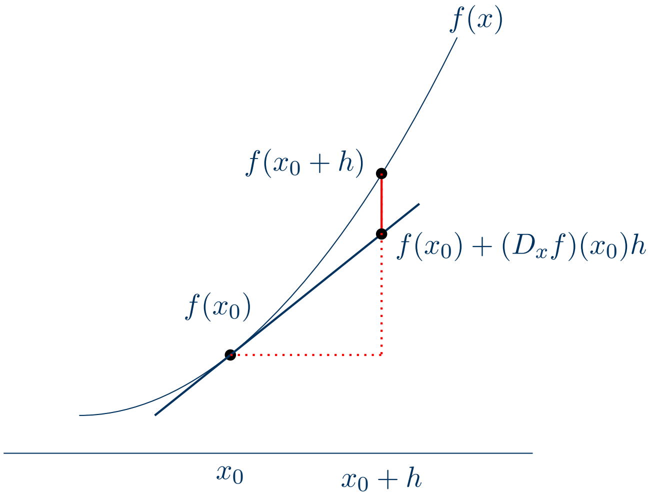
Derivative as linear approximation



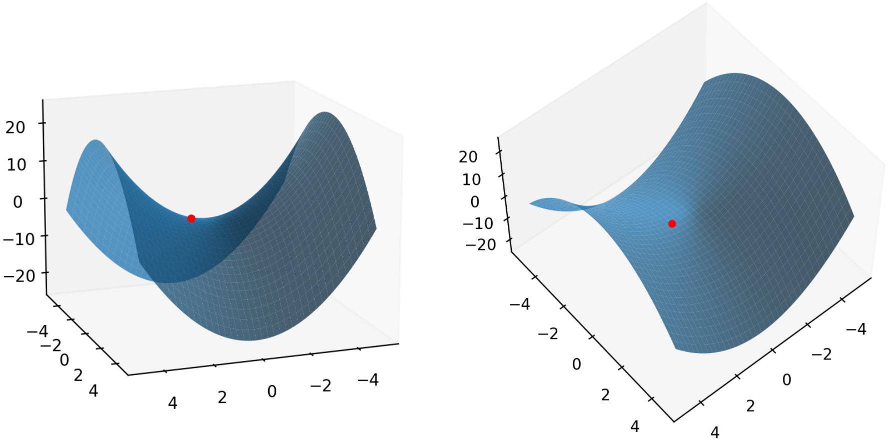
Derivative as linear approximation



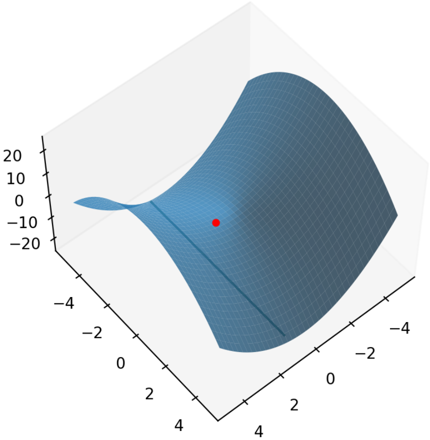
Derivative as linear approximation

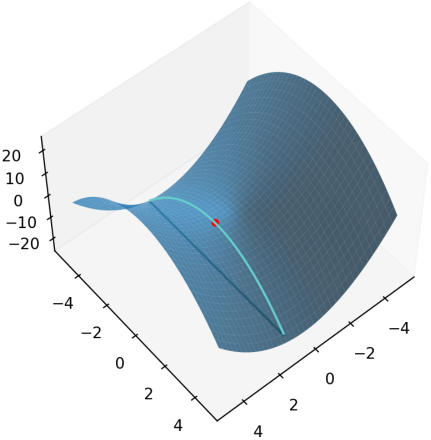


Derivative in 3D

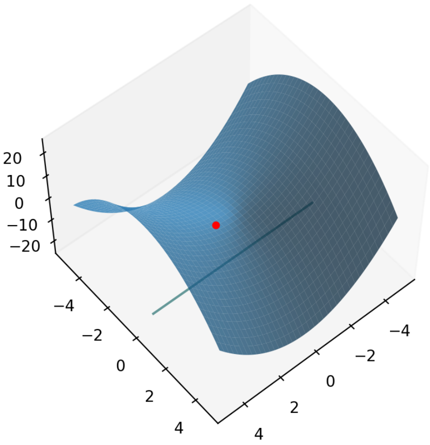


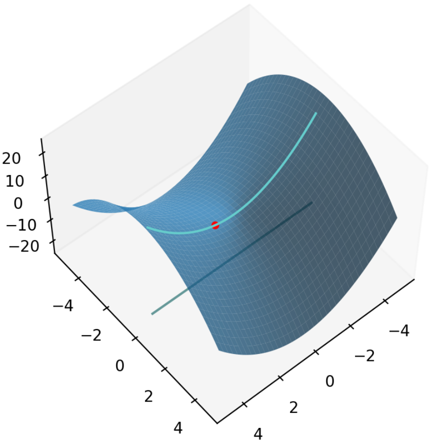
Derivative in 3D





Derivative in 3D





Directional derivative

- The directional derivative of $f : \mathbb{R}^d \to \mathbb{R}$ along the direction v at $x_0 \in \mathbb{R}^d$ is defined as

$$(D_v f)(x_0) = \lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t}.$$

Directional derivative

- The directional derivative of $f : \mathbb{R}^d \to \mathbb{R}$ along the direction v at $x_0 \in \mathbb{R}^d$ is defined as

$$(D_v f)(x_0) = \lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t}.$$

- If we let $g(t) = f(x_0 + tv)$, then

$$\begin{aligned}(D_v f)(x_0) = \lim_{t \to 0} \frac{f(x_0 + tv) - f(x_0)}{t} &= \lim_{t \to 0} \frac{g(t) - g(0)}{t} \\ &= \lim_{t \to 0} \frac{g(0 + t) - g(0)}{t} = (D_t g)(0)\end{aligned}$$

Example

- Consider the function $f(x, y) = x^2 - y^2$.

- If we are at $(2, 0)$, the directional derivative along $(1, 0)$ is 4.

- If we take a line at $\{(x, y) : (x, y) = (2, 0) + t(1, 0) = (2 + t, 0) \text{ for } t \in \mathbb{R}\}$, we have $g(t) = f(2 + t, 0) = (2 + t)^2$. The derivative $(D_t g)(t) = 2(2 + t)$, and $(D_t g)(0) = 2 \cdot (2 + 0) = 4$.

Partial derivatives

- A partial derivative is a directional derivative along the direction of coordinate axes.
- In a three-dimensional space, the direction of the axes are

$$(1,0,0) \quad (0,1,0) \quad (0,0,1).$$

For a function $f : \mathbb{R}^3 \to \mathbb{R}$, the partial derivatives along the axes are

$$\frac{\partial}{\partial x} f \quad \frac{\partial}{\partial y} f \quad \frac{\partial}{\partial z} f.$$

Example

- Given a function $f(x, y) = x^2 - y^2$, show that

$$\left(\frac{\partial}{\partial x} f\right)(x, y) = 2x \qquad \left(\frac{\partial}{\partial y} f\right)(x, y) = -2y.$$

- The x-axis is the direction $(1, 0)$. At any point (x, y), the line along that direction is $(x + t, y)$. The function value along that line is $g(t) = f(x + t, y) = (x + t)^2 - y^2$. We then have $(D_t g)(t) = 2(x + t)$, and

$$\left(\frac{\partial}{\partial x} f\right)(x, y) = (D_t g)(0) = 2x.$$

- Treat other variables as constants and take 1D derivatives.

Example

- Given a function $f(x,y,z) = (x+2y-3z)^2$, show that

$$\left(\frac{\partial}{\partial x} f\right)(x,y,z) = 2(x+2y-3z)$$
$$\left(\frac{\partial}{\partial y} f\right)(x,y,z) = 2(x+2y-3z)\cdot 2$$
$$\left(\frac{\partial}{\partial z} f\right)(x,y,z) = 2(x+2y-3z)\cdot(-3)$$

Example

- Given a sigmoid function

$$f(w, b) = \frac{1}{1 + \exp(-(w^\top x + b))},$$

show that

$$\left(\frac{\partial}{\partial b} f \right)(w, b) = f(w, b)(1 - f(w, b))$$

Gradient

- The gradient of a function is the vector consisting of all partial derivatives.
- For a function $f : \mathbb{R}^3 \to \mathbb{R}$, its gradient is

$$(\nabla f)(x, y, z) = \begin{bmatrix} \left(\frac{\partial}{\partial x} f\right)(x, y, z) \\ \left(\frac{\partial}{\partial y} f\right)(x, y, z) \\ \left(\frac{\partial}{\partial z} f\right)(x, y, z) \end{bmatrix}.$$

Example

- Given a function $f(x, y, z) = (x + 2y - 3z)^2$, show that its gradient is

$$(\nabla f)(x, y, z) = \begin{bmatrix} 2(x + 2y - 3z) \\ 2(x + 2y - 3z) \cdot 2 \\ 2(x + 2y - 3z) \cdot (-3) \end{bmatrix}.$$

Example

- Given a function $f(a) = b^\top a$, show that its gradient is

$$(\nabla f)(a) = b.$$

- Given a function $f(a) = b^\top A a$, show that its gradient is

$$(\nabla f)(a) = A^\top b.$$

- Given a function $f(a) = \|a\|_2^2$, show that its gradient is

$$(\nabla f)(a) = 2a.$$

Example

- Given a function $f(w) = (w^\top x + b - y)^2$, show that

$$(\nabla f)(w) = 2(w^\top x + b - y)x.$$

- Given a function $f(w) = \dfrac{1}{1 + \exp(-(w^\top x + b))}$ show that its gradient is

$$(\nabla f)(w) = f(w)(1 - f(w))x.$$

Theorem

- For a function $f : \mathbb{R}^d \to \mathbb{R}$ and any direction v at any point x, show that

$$(D_v f)(x) = (\nabla f)(x)^\top v.$$

- Once we know the gradient, we know all directional derivatives.

Second-order derivative

For a function $f : \mathbb{R} \to \mathbb{R}$, its second-order derivative is defined and written as

$$\frac{\partial^2}{\partial x^2} f = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial x} f\right).$$

- Given a function $f(x) = x^2$, it's second-order derivative is 2.
- The second-order derivative tells us whether the function looks like a cup or an upside-down cup.

Hessian

- The Hessian of a function $f : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\begin{bmatrix} \frac{\partial^2}{\partial x_1 \partial x_1} f & \frac{\partial^2}{\partial x_1 \partial x_2} f & \cdots & \frac{\partial^2}{\partial x_1 \partial x_d} f \\ \frac{\partial^2}{\partial x_2 \partial x_1} f & \frac{\partial^2}{\partial x_2 \partial x_2} f & \cdots & \frac{\partial^2}{\partial x_2 \partial x_d} f \\ \vdots f & & & \vdots \\ \frac{\partial^2}{\partial x_d \partial x_1} f & \frac{\partial^2}{\partial x_d \partial x_2} f & \cdots & \frac{\partial^2}{\partial x_d \partial x_d} f \end{bmatrix}.$$

- The Hessian matrix is always symmetric, because

$$\frac{\partial^2}{\partial x_j \partial x_i} f = \frac{\partial^2}{\partial x_i \partial x_j} f,$$

- Given a function $f(x, y) = x^2 - y^2$, show that its Hessian is $\begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$.