Calculus refresher

Statistical Natural Language Processing 1

Çağrı Çöltekin

University of Tübingen
Seminar für Sprachwissenschaft

Winter Semester 2025/2026

version: 81f9cf5 @2025-12-03

Calculus and NLP

- Our main use case is finding minima:
  - Supervised ML models are (typically) trained by minimizing an error, or a *loss function*
  - Differential calculus allows efficiently searching minima, and determining where minima are
- Integrals will also be handy for calculating probabilities

Today's plan

Very brief introductions to

- Limits
  - We are mainly interested in for defining derivatives
- Derivatives
  - Are the central topic for us: training a ML system relies on derivation (but there are also other interesting uses)
- Integrals
  - Mainly for probability theory, but without integrals, the derivatives would also be incomplete

Limits

- Limit is the value that a function approaches as its argument approaches (arbitrarily close, but not equal) to some value.
- We write
$$\lim_{x \to c} f(x)$$
for the value of function f as x approaches to c.
- If the value of the function at $x = c$ is a number, the limit is $f(c)$
$$\lim_{x \to 2} x^2 = 2^2 = 4$$
- If the result is ∞, the limit does not exist
- Interesting cases are when the function is discontinuous, or undefined, e.g., $f(x)$ is $0/0$, ∞/∞, $\infty - \infty$

Example (1)

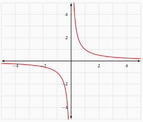
$$\lim_{x \to 1} \frac{1}{x} = 1$$

$$\lim_{x \to 0} \frac{1}{x} = \pm\infty$$

More precisely,

$$\lim_{x \to 0^-} \frac{1}{x} = -\infty$$

$$\lim_{x \to 0^+} \frac{1}{x} = \infty$$



Example (2)

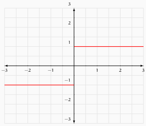
$$f(x) = \begin{cases} -1 & x < 0 \\ 1 & x > 0 \end{cases}$$

$$\lim_{x \to 0} f(x) = ?$$

$$\lim_{x \to 0^-} f(x) = -1$$

$$\lim_{x \to 0^+} f(x) = 1$$

limit does not exist

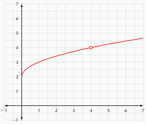


Example (3)

$$\lim_{x \to 4} \frac{x-4}{\sqrt{x}-2}$$

$$\lim_{x \to 4^-} \frac{x-4}{\sqrt{x}-2} = 4$$

$$\lim_{x \to 4^+} \frac{x-4}{\sqrt{x}-2} = 4$$



Rules for limits

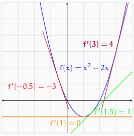
- $\lim_{x \to c} (f(x) + g(x)) = \lim_{x \to c} f(x) + \lim_{x \to c} g(x)$
- $\lim_{x \to c} af(x) = a \lim_{x \to c} f(x)$
- $\lim_{x \to c} (f(x)g(x)) = \lim_{x \to c} f(x) \lim_{x \to c} g(x)$

Derivatives

- Derivative of a function $f(x)$ is another function $f'(x)$ indicating the rate of change in $f(x)$
- Alternative notation: $f'(x) = \frac{df}{dx}(x)$
- When derivative exists, it determines the tangent line to the function at a given point
- Example from physics: velocity is the derivative of the position
- Our main interest:
  - the points where the derivative is 0 are the stationary points (maxima, minima, inflection points)
  - the derivative evaluated at other points indicate the direction and steepness of the curve defined by the function

Example: derivatives

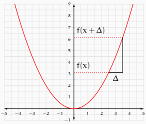
- $f'(x)$ is negative when $f(x)$ is decreasing, positive when it is increasing
- The absolute value of $f'(x)$ indicates how fast $f(x)$ changes when x changes
- $f'(x) = 0$ when at a *stationary point*
- $f'(a)$ is a (good) approximation to the $f(x)$ near the a



Derivative of a function

definition

$$f'(x) = \lim_{\Delta \to 0} \frac{f(x+\Delta) - f(x)}{\Delta}$$



Example: calculating derivatives using the definition

$$\begin{aligned} \frac{d}{dx}x^2 &= \lim_{\Delta \to 0} \frac{(x+\Delta)^2 - x^2}{\Delta} \\ &= \lim_{\Delta \to 0} \frac{x^2 + 2\Delta x + \Delta^2 - x^2}{\Delta} \\ &= 2x \end{aligned}$$

Some derivatives to know

- Powers: $\frac{d}{dx}x^n = nx^{n-1}$
- Trigonometric functions:
  $\frac{d}{dx}\sin(x) = \cos(x)$
  $\frac{d}{dx}\cos(x) = -\sin(x)$
  $\frac{d}{dx}\tan(x) = 1 + \tan^2(x)$
- Powers of e: $\frac{d}{dx}e^x = e^x$
- Natural logarithm: $\frac{d}{dx}\ln(x) = \frac{1}{x}$

General rules for derivatives

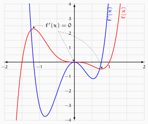
- Sum rule:
$$\frac{d}{dx}f(x) + g(x) = \frac{d}{dx}f(x) + \frac{d}{dx}g(x)$$
- Product rule:
$$\frac{d}{dx}f(x)g(x) = g(x)\frac{d}{dx}f(x) + f(x)\frac{d}{dx}g(x)$$
- Chain rule: if $f(x) = h(g(x))$
$$\frac{df}{dx} = \frac{dh}{dg}\frac{dg}{dx} \quad \text{, or} \quad f'(x) = h'(g(x))g'(x)$$

Cain rule: examples

$$\frac{d}{dx}e^{x^3} \qquad\qquad \frac{d}{dx}2^{\sin x}$$

Derivatives and extrema

- Derivative of a function is 0 at minimum, maximum and inflection points
- Derivative is useful for optimization (minimization of maximization) problems
- We need additional tests to determine the type of critical points

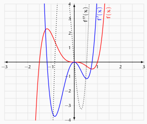


Higher order derivatives

- Higher order derivatives, particularly the second derivative, are useful in many applications
  - Determining the type of critical points
  - Polynomial approximations to functions
- Notation:
  - Second derivative: $f''(x) = \frac{d^2f}{dx^2}$
  - n^{th} derivative: $f^{(n)}(x) = \frac{d^nf}{dx^n}$

Second derivatives and extrema

- Second derivatives are useful for determining the type of critical points
  - $f''(x) < 0$ if $f(x)$ is concave down ($\cap$)
  - $f''(x) > 0$ if $f(x)$ is concave up ($\cup$)
  - $f''(x) = 0$ if $f(x)$ is flat

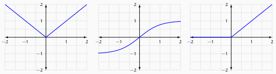


Differentiable functions and continuity

- A function is said to be *differentiable* if its derivative exists at every point in its domain
- This concept is important when we want to use optimization techniques based on derivatives
- A differentiable function is also continuous, but a continuous function is not necessarily differentiable

Differentiability

Are these functions differentiable?



Partial derivatives and gradient

- In ML, we are often interested in (error) functions of many variables
- A partial derivative is the derivative of a multivariate function with respect to a single variable while treating all others as constants. For $f(x, y)$,
  $\frac{\partial f}{\partial x}$ is the partial derivative with respect to x
  $\frac{\partial f}{\partial y}$ is the partial derivative with respect to y
- A very useful quantity, called *gradient*, is the vector of partial derivatives with respect to each variable

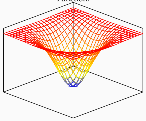
$$\nabla f(x_1, \ldots, x_n) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)$$

- Gradient points to the direction of the steepest increase
- Example: if $f(x, y) = x^3 + yx$

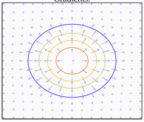
$$\nabla f(x, y) = \left(3x^2 + y, x\right)$$

Gradient visualization

Function:



Gradients:

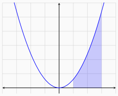


Integrals

- Integral is the reverse of the derivative (anti-derivative)
- The indefinite integral of $f(x)$ is noted $F(x) = \int f(x)dx$
- We are often interested in definite integrals

$$\int_a^b f(x)dx = F(b) - F(a).$$

- Integral gives the area under the curve



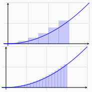
Numeric integrals & infinite sums

- The integral can be defined as

$$\int_a^b f(x)dx = \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i)\Delta x$$

where $\Delta x = \frac{b-a}{n}$

- Integration is 'infinite summation'
- As the width of the rectangles converges to 0 (or number of rectangles becomes ∞), the sum converges to the area under the curve
- When integration is not possible with analytic methods, we resort to numeric integration



Summary / next

We reviewed three main concepts from calculus

- Limits
- Derivatives
- Integrals

Next:

- Regression again: through gradient optimization
- Introduction to probability theory

Further reading

- A nice video series: `https://www.youtube.com/playlist?list=PLZHQObOWTQDMsr9K-rj53DwVRMYO3t5Yr`
- No concrete reading suggestions, but check `https://www.openculture.com/free-math-textbooks`