# Calculus

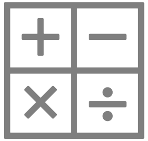
Holberton

# Why Calculus in Machine Learning

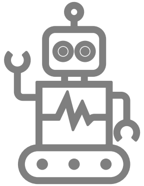
# Why Calculus for Machine Learning

- Framework for understanding complex systems

- Theoretical foundation for ML algorithms
  - E.g. gradient descent optimization; backpropagation



- Optimize ML algorithms
  - E.g. loss function (accuracy of prediction)



- Calculate derivatives and integrals
  - Understand functions and data

# Mathematical Series

# Mathematical Series $\sum \quad \prod$

- Mathematical series = sum of a sequence of terms

- Sigma notation ( $\sum$ ) = summation of a series of terms
  - Example: $\Sigma_{i=1}^{10} i = 1 + 2 + \cdots + 10 = 55$

- Product notation ( $\prod$ ) = product of a series of terms
  - Example: $\Pi_{i=1}^{10} i = 1 \cdot 2 \cdot \ldots \cdot 10 = 3628800$

# Mathematical Series

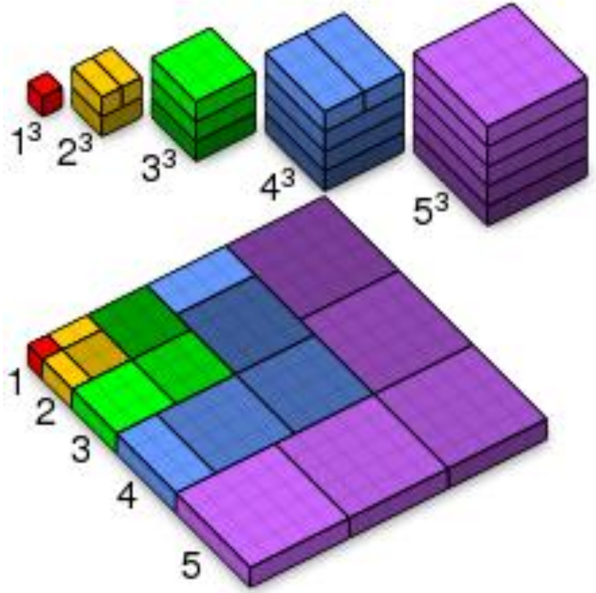
- Mathematical series = sum of a sequence of terms

- Sigma notation ( $\sum$ ) = summation of a series of terms
  - Example: $\Sigma_{i=1}^{10} i = 1 + 2 + \cdots 10 = 55$
  - Example: $\Sigma_{i=1}^{3} i^2 = 1^2 + 2^2 + 3^2 = 14$
- Product notation ( $\prod$ ) = product of a series of terms
  - Example: $\Pi_{i=1}^{10} i = 1 \cdot 2 \cdot \ldots \cdot 10 = 3628800$
  - Example: $\Pi_{i=1}^{2} \Pi_{j=4}^{6} (3 \cdot i \cdot j) = \Pi_{i=1}^{2} \left( (3i \cdot 4) \cdot (3i \cdot 5) \cdot (3i \cdot 6) \right) =$ $\left( (3 \cdot 1 \cdot 4)(3 \cdot 1 \cdot 5)(3 \cdot 1 \cdot 6) \right) \left( (3 \cdot 2 \cdot 4)(3 \cdot 2 \cdot 5)(3 \cdot 2 \cdot 6) \right) =$

# Famous Mathematical Series

- Faulhaber‘s formula
  - $\Sigma_{k=1}^{\mathrm{n}} k^p = 1^p + 2^p + \cdots k^p$
  - $\Sigma_{k=1}^{m} k = \frac{m\,(m+1)}{2}$
  - $\Sigma_{k=1}^{m} k^2 = \frac{m\,(m+1)(2m+1)}{6} = \frac{m^3}{3} + \frac{m^2}{2} + \frac{m}{6}$
  - $\Sigma_{k=1}^{m} k^3 = \left[\frac{m\,(m+1)}{2}\right]^2 = \frac{m^4}{4} + \frac{m^3}{2} + \frac{m^2}{4}$

Proof without words



# Bernoulli numbers

- A sequence of rational numbers which occur frequently in data analysis

- Useful in modeling data and approximating functions

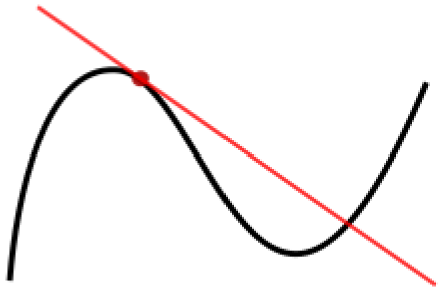
- **Bernoulli polynomials:** used for series expansion & approximation of high-dimensional polynomials

$$\frac{x}{e^x - 1} \equiv \sum_{n=0}^{\infty} \frac{B_n x^n}{n!}.$$

# Derivatives

# Derivatives

$\Delta$ $\frac{\delta y}{\delta x}$



- Derivative = rate of change

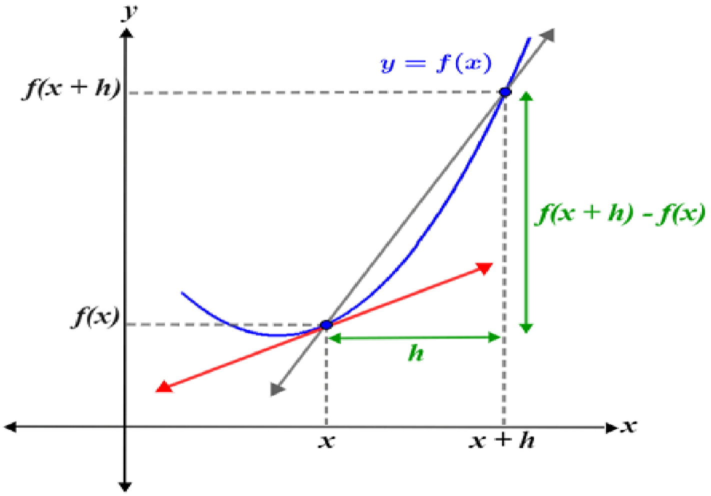
- Describes by “how much„ will a function change when its input will change?

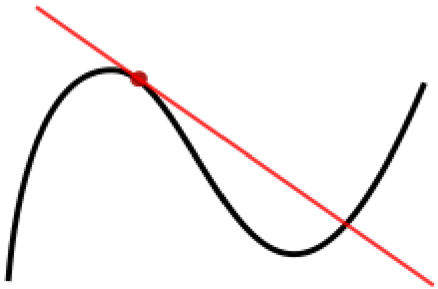

- ML application: adjust parameters to minimize “loss„ (e.g., gradient descent or chain rule)

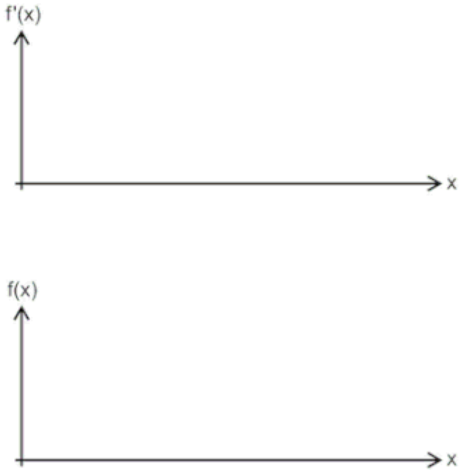
# Derivatives

- Derivative = rate of change

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



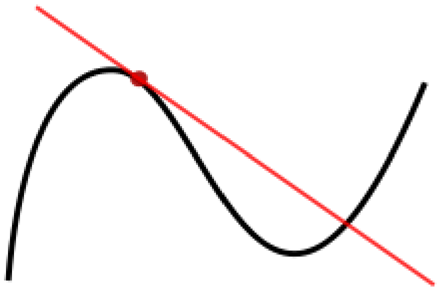


# Derivatives of functions

- Linear functions $(a \cdot x + b)$ → **constant**

- Power functions $(x^a)$
  - $\frac{d}{dx}\, x^a = a \cdot x^{a-1}$
  - $\frac{d}{dx}\left(\frac{1}{x}\right) = \left(\frac{-1}{x^2}\right)$

- Logarithmic functions
  - $\frac{d}{dx}\ln(x) = \frac{1}{x}$




# Rules of derivatives

- **Sum rule** of derivatives
  - $h(x) = f(x) + g(x) \rightarrow \; h'(x) = f'(x) + g'(x)$
  - ML example: derivative of cost function is the sum of individual loss functions for each data point

- **Product rule** of derivatives
  - $h(x) = f(x) \cdot g(x) \rightarrow \; h'(x) = f'(x) \cdot g(x) + g'(x) \cdot f(x)$
  - ML example: optimization process (taking derivative of combined functions)

# Rules of derivatives

- **Chain rule** of derivatives
  - Derivative of a composite function, e.g. $f(g(x))$
  - Take into account the effect of both functions
  - ML application: compute gradients of a neural network (the backpropagation algorithm)
- Formula
  - Let $u = g(x)$ and $y = f(u)$, then $\frac{\delta y}{\delta x} = \frac{\delta y}{\delta u} \cdot \frac{\delta u}{\delta x} = f'(u) \cdot g'(x)$
  - In Leibniz notation: $\frac{\delta}{dx} f\big(g(x)\big) = f'\big(g(x)\big) \cdot g'(x)$

# Integrals

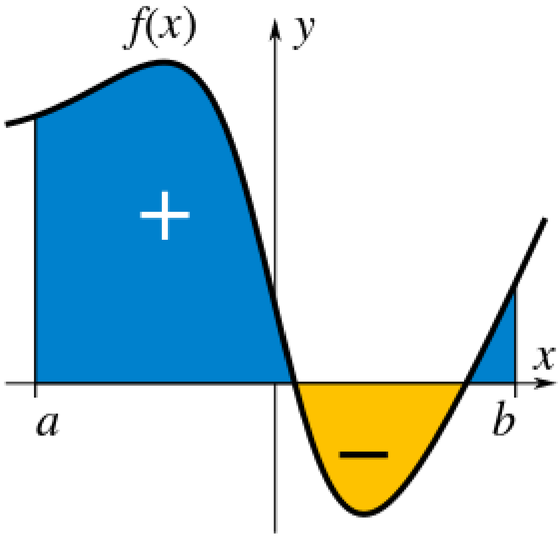
# Integrals

$$\int_1^n x$$

- Derivative = contigious analog of a sum

- Find area under a curve

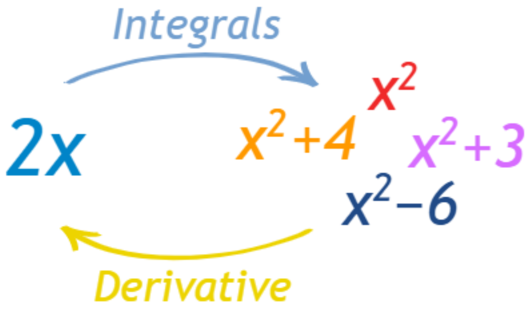
- ML application: probability density functions or optimization

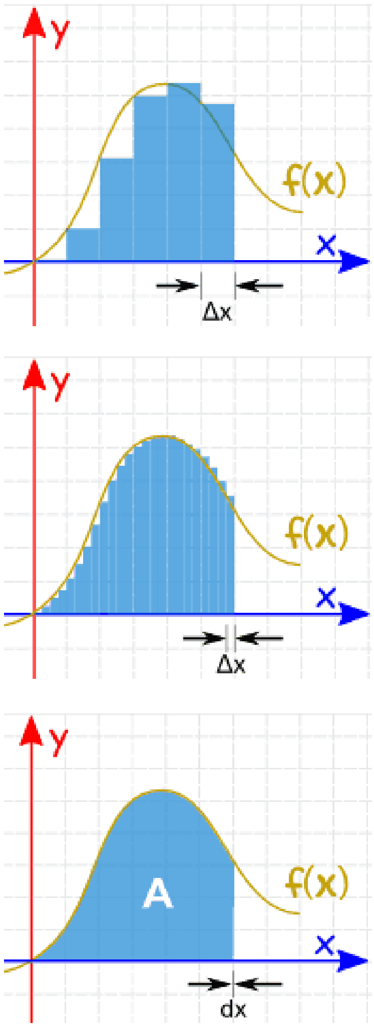


## Integrals

$$\int_{1}^{n} x$$

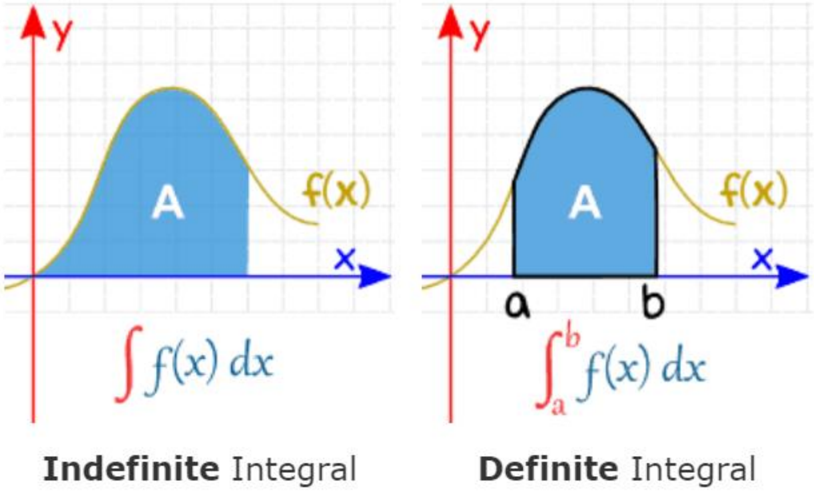
- Integrals are the **reverse** of derivatives



# Integrals

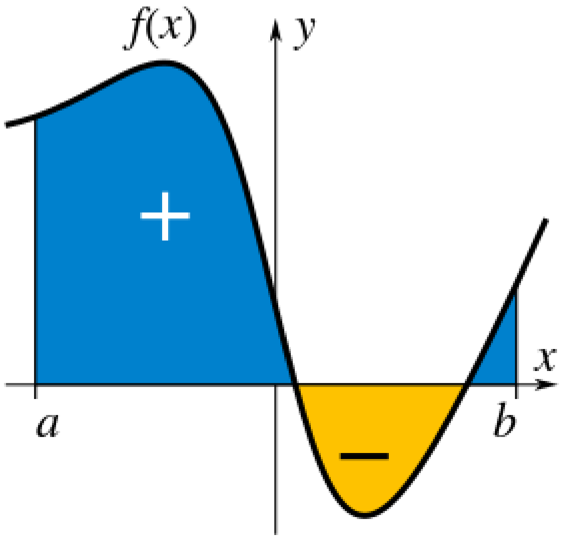
- Definite **vs** indefinite integrals




- Definite integrals have **actual values** to calculate within

# Integrals: basic formulas

- $\int 1\, dx = x + C$
- $\int a\, dx = ax + C$
- $\int x^n\, dx = \frac{x^{n+1}}{n+1} + C$
- $\int \frac{1}{x} dx = \ln|x| + C$
- $\int a^x\, dx = \frac{a^x}{\ln a} + C$



# Any questions?