

ML ERASMUS+, NOV 28, 2022

---

## Gradients and optimization

$$\frac{\partial C(\beta)}{\partial \beta} = 0 = g(\beta)$$

Newton's method

$$\beta^{(n+1)} = \beta^{(n)} - H^{-1}(\beta^{(n)}) g(\beta^{(n)})$$

$$\approx \beta^{(n)} - \eta^{(n)} g(\beta^{(n)})$$



- fixed learning rate
- scheduler for  $\eta^{(n)}$ 
  - linear
  - exponential
- Adaptive (with info on  $g(\beta^{(n)})$ )
  - Adagrad
  - RMS prop
  - ADAM

SGD/GD  
with  
momentum

— Automatic differentiation

— Difference methods—

$$\frac{df(x)}{dx} \approx \frac{f(x+\Delta x) - f(x-\Delta x)}{2\Delta x}$$

$(O(\Delta x^2))$

— Symbolic manipulations—

— Analytical derivations—