Basics: Continuous Optimization

Terminology:

$$\left.\begin{aligned} f &: \mathbb{R} \to \mathbb{R} \\ f &: \mathbb{R}^d \to \mathbb{R} \\ f &: \mathbb{R}^{m \times n} \to \mathbb{R} \end{aligned}\right\} \text{objective function}$$

$f: \mathbb{R} \to \mathbb{R}$

Optimization problem:

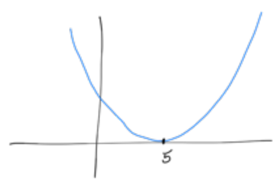
$$f^* = \min_{x \in \mathbb{R}} f(x)$$

$$f^* = \max_{x \in \mathbb{R}} f(x)$$

$$x^* = \operatorname{argmin}_{x \in \mathbb{R}} f(x)$$

$$x^* = \operatorname{argmax}_{x \in \mathbb{R}} f(x)$$

Ex: $f(x) = (x-5)^2$: minimization



$$f^* = \min_{x \in \mathbb{R}} f(x) = 0$$

$$x^* = \operatorname{argmin}_{x \in \mathbb{R}} f(x) = 5$$

Gradient Descent

(iterative process) "Gradient"

x^0 $\quad$ x^1 $\quad \nabla f(x^0) < 0 \Rightarrow x^1 > x^0$

$\quad\quad\quad\quad\quad\;\; \nabla f(x^0) > 0 \Rightarrow x^1 < x^0$

η: "step size" > 0 $\quad \eta \in \mathbb{R}$

$$x^1 = x^0 - \eta \nabla f(x^0)$$

$$\boxed{x^k = x^{k-1} - \eta \nabla f(x^{k-1})}$$

"Convergence" at x^*: $\nabla f(x^*) = 0$

Apply GD: $f^* = \min_{x \in \mathbb{R}} (x-5)^2$

$x^0 = 7 \quad \eta = 0.25$

$f(x) = (x-5)^2$

$\nabla f(x) = 2(x-5)$

① $x^1 = x^0 - \eta \nabla f(x^0)$
$= 7 - 0.25 \times 2 \times 2$
$= 7 - 1 = 6$

② $x^2 = x^1 - \eta \nabla f(x^1)$
$= 6 - 0.25 \times 2 \times 1$
$= 6 - 0.5 = 5.5$

③ $x^3 = x^2 - \eta \nabla f(x^2)$
$= 5.5 - 0.25 \times 2 \times 0.5$
$= 5.5 - 0.25 = 5.25$

$\vdots$

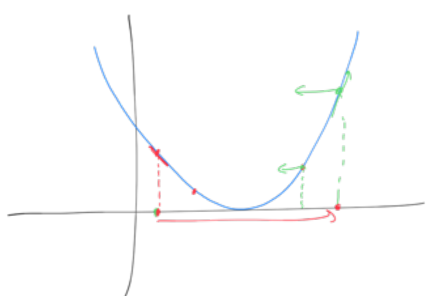
Effect of step-size

① $\eta = 1$

$x^0 = 7$

$x^1 = x^0 - \eta \nabla f(x^0)$
$= 7 - 1 \times 2 \times 2$
$= 3$

② $x^2 = x^1 - \eta \nabla f(x^1)$
$= 3 - 1 \times 2 \times (-2)$
$= 3 + 4 = 7$



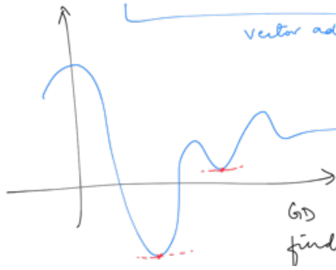
① Step size too large: oscillations (no convergence)

② Step size too small: slow convergence

$f: \mathbb{R}^d \to \mathbb{R}$

$x \in \mathbb{R}^d$

$$x^{k} = \underbrace{x^{k-1} - \underbrace{\overset{\text{scalar}}{\eta}\ \underbrace{\nabla f(x^{k-1})}_{\text{vector}}}_{\text{scalar multiplication}}}_{\text{vector addition}}$$



GD won't always find the optimal solution