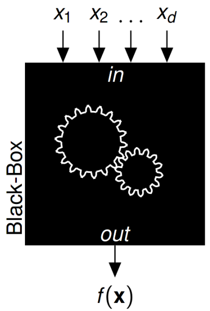
# Optimization in Machine Learning

# Bayesian Optimization
# Black Box Optimization




**Learning goals**

- Definition and properties
- Examples
- Naive approaches

# STANDARD VS. BLACK-BOX OPTIMIZATION

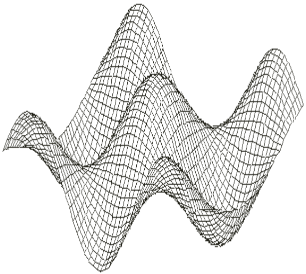
**Optimization:** Find

$$\min_{\mathbf{x} \in \mathcal{S}} f(\mathbf{x})$$

with objective function

$$f : \mathcal{S} \rightarrow \mathbb{R},$$

where $\mathcal{S}$ is usually box constrained.



If we are lucky ...

- ... we have an analytic description of $f : \mathcal{S} \rightarrow \mathbb{R}$
- ... we can calculate gradients and use gradient-based methods (e.g. gradient descent) for optimization

# STANDARD VS. BLACK-BOX OPTIMIZATION / 2

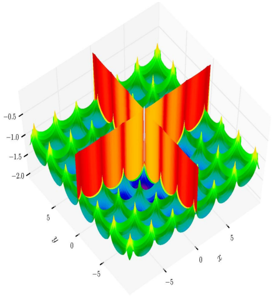
**Optimization:** Find

$$\min_{\mathbf{x} \in \mathcal{S}} f(\mathbf{x})$$

with objective function

$$f: \ \mathcal{S} \rightarrow \mathbb{R},$$

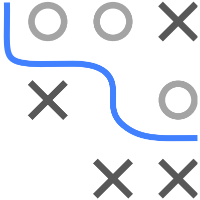
where $\mathcal{S}$ is usually box constrained.




Optimization gets harder ...

- ... if we cannot calculate gradients (because $f$ is not differentiable or $f$ is not known to us)
- ... but as long as evaluations of $f$ are cheap, we can use standard derivative-free optimization methods (e.g. Nelder-Mead, simulated annealing, EAs)

# STANDARD VS. BLACK-BOX OPTIMIZATION / 3



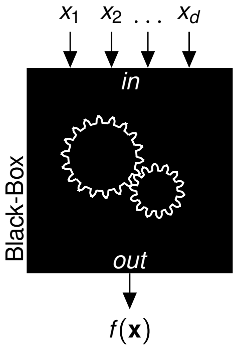
**Optimization:** Find

$$\min_{\mathbf{x} \in \mathcal{S}} f(\mathbf{x})$$

with objective function

$$f : \mathcal{S} \to \mathbb{R},$$

where $\mathcal{S}$ is usually box constrained.

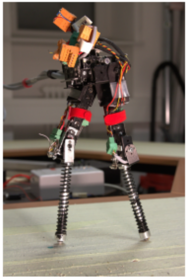



Optimization gets **really hard** if ...

- ... there is no analytic description of $f : \mathcal{S} \to \mathbb{R}$ (**black box**)
- ... evaluations of $f$ for given values of $\mathbf{x}$ are **time consuming**

# EXAMPLES FOR BAYESIAN OPTIMIZATION

1. Robot Gait Optimization: The robot's gait is controlled by a **parameterized controller**



- **Goal:** Find parameters s.t. average velocity (directional speed) of the robot is maximized
- Parameters of the gait control e.g. joints of ankles and knees
- *Calandra et al. (2014). An Experimental Evaluation of Bayesian Optimization on Bipedal Locomotion*

# EXAMPLES FOR BAYESIAN OPTIMIZATION / 2

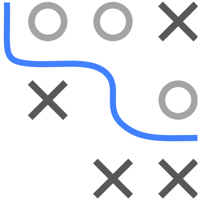
❷ Optimization of a cookie recipe

https://www.bettycrocker.com

| Ingredient | Salt (tsp)† | Total Sugar (g) | Brown Sugar (%) | Vanilla (tsp)† | Chip Quantity (g) | Chip Type |
|---|---|---|---|---|---|---|
| Min | 0 | 150 | 0 | 0.25 | 114 | {Dark, Milk, |
| Max | 0.5 | 500 | 1 | 1 | 228 | White} |

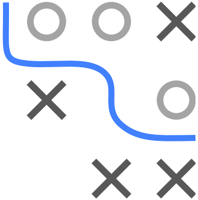
- **Goal:** Find “optimal” composition and amounts of ingredients
- **Evaluation:** Cookies are baked according to the recipe, tested and rated by volunteers
- *Kochanski et al. (2017). Bayesian Optimization for a Better Dessert*

# NAIVE APPROACHES

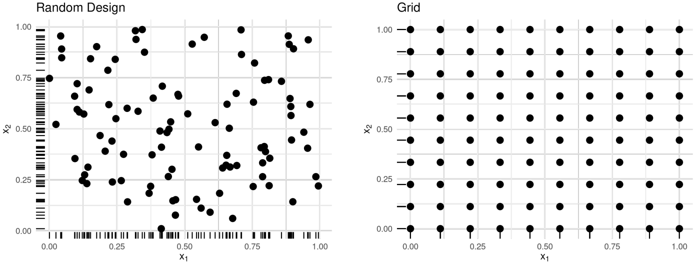


1. Empirical knowledge / manual tuning
   - Select parameters based on “expert” knowledge
   - **Advantages:** Can lead to fairly good outcomes for known problems
   - **Disadvantages:** Very (!) inefficient, poor reproducibility, chosen solution can also be far away from a global optimum

# NAIVE APPROACHES / 2



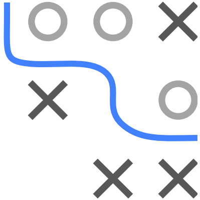
2. Grid search / random search
   - Grid search: Exhaustive search of a predefined grid of inputs
   - Random search: Evaluate uniformly sampled inputs
   - **Advantages:** Easy, intuitive, parallelization is trivial
   - **Disadvantages:** Inefficient, search large irrelevant areas



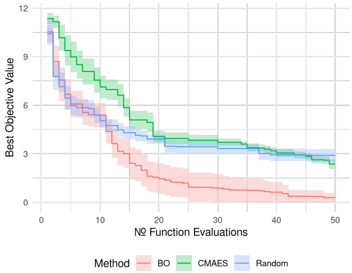

Rug plots of RS vs. GS.

# NAIVE APPROACHES / 3

3. Traditional black-box optimization
   - Traditional approaches that do not require derivatives
   - E.g. Nelder-Mead, simulated annealing, EAs
   - **Advantages:** Truly iterative, focuses on relevant regions
   - **Disadvantages:** Still inefficient; usually lots of evaluations are needed to produce good outcomes



# NAIVE APPROACHES / 4




BO vs. CMAES vs. RS on 2D Ackley.