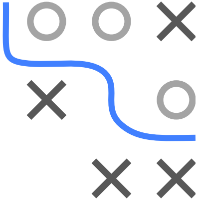
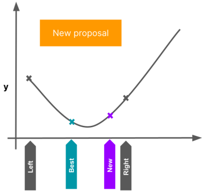
Optimization in Machine Learning



Univariate optimization
Golden ratio



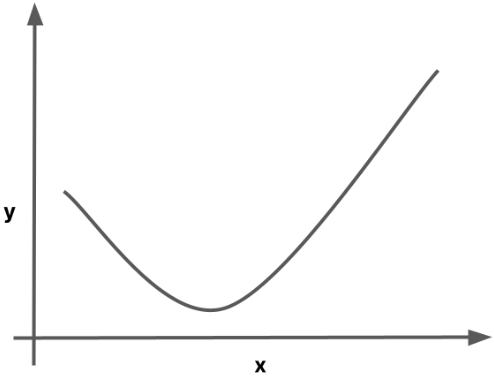
Learning goals

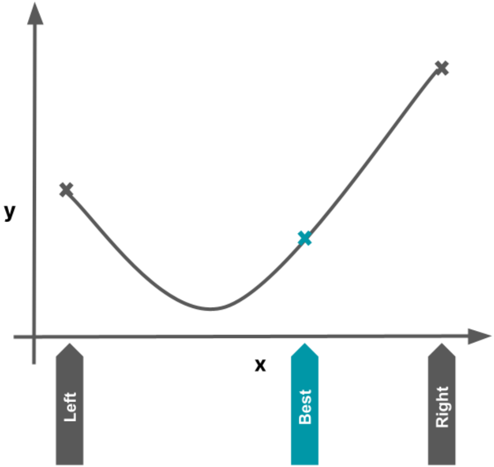
- Simple nesting procedure
- Golden ratio

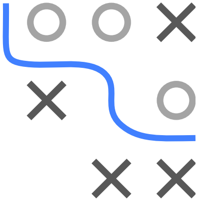
UNIVARIATE OPTIMIZATION

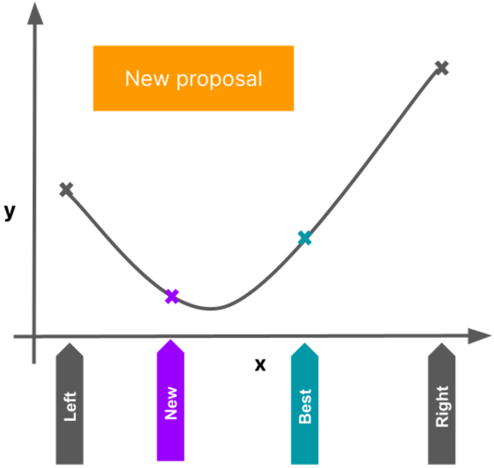
Let $f : \mathbb{R} \to \mathbb{R}$ (search over interval $(x_{\text{left}}, x_{\text{right}})$ in practice)

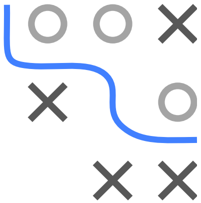
Goal: Iteratively improve eval points. Assume function $-f$ is unimodal.
Will not rely on gradients, so this also works for black-box problems.

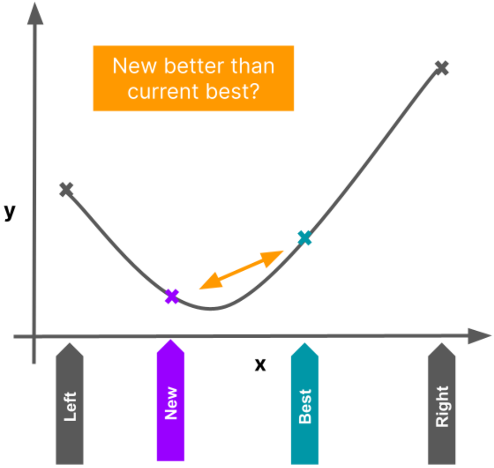


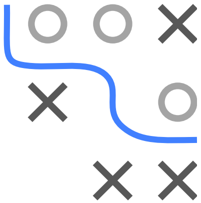

y
x
Left
Best
Right

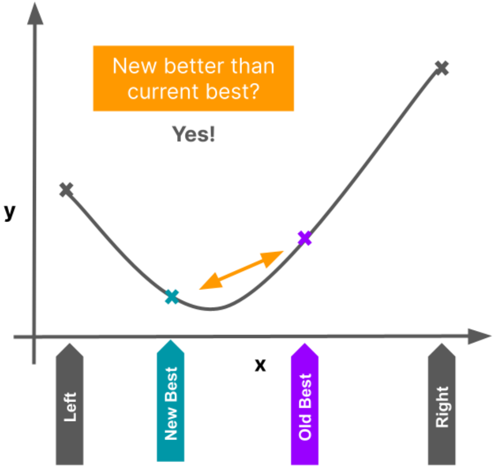



New proposal
y
x
Left
New
Best
Right

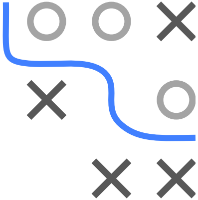


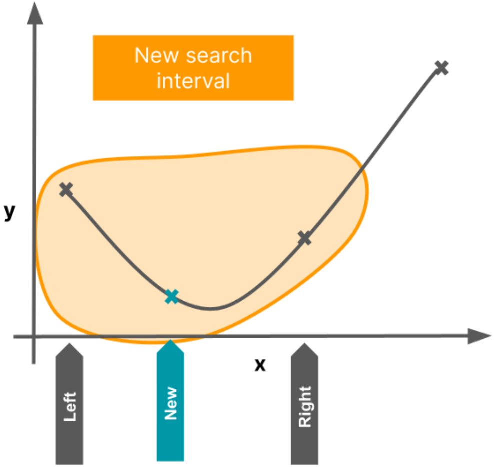

New better than current best?
y
x
Left
New
Best
Right

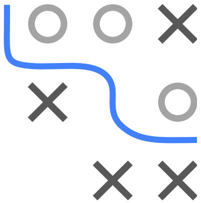


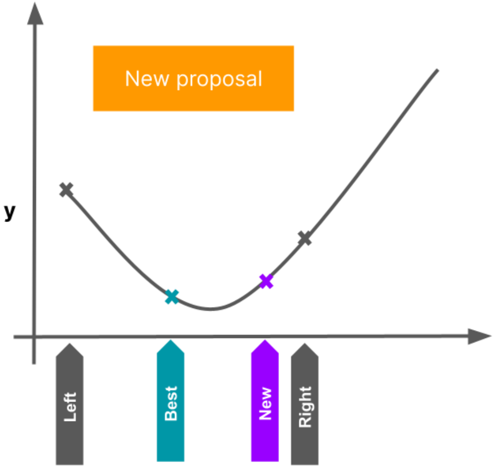


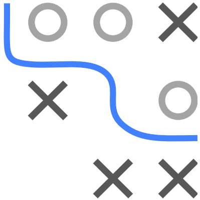
New better than current best?
Yes!
y
x
Left
New Best
Old Best
Right

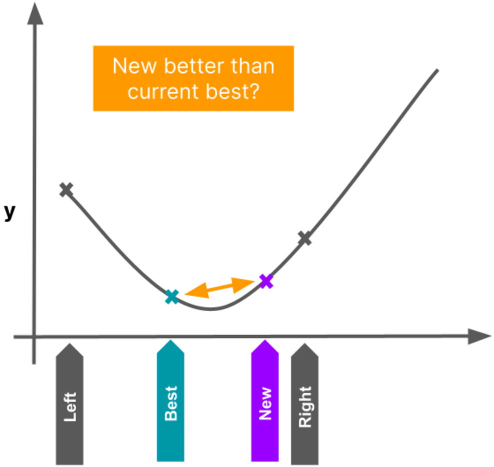


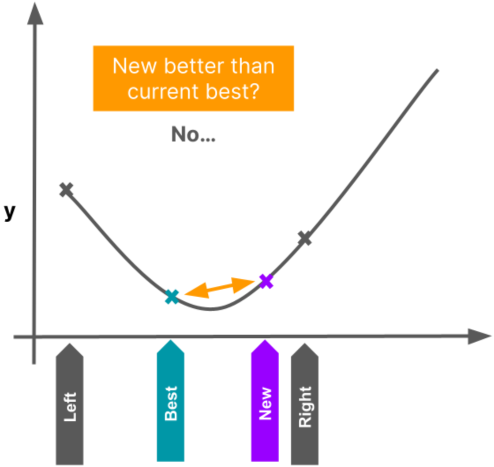

New search interval
y
x
Left
New
Right

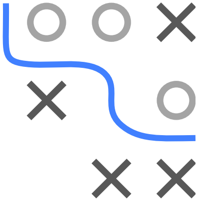


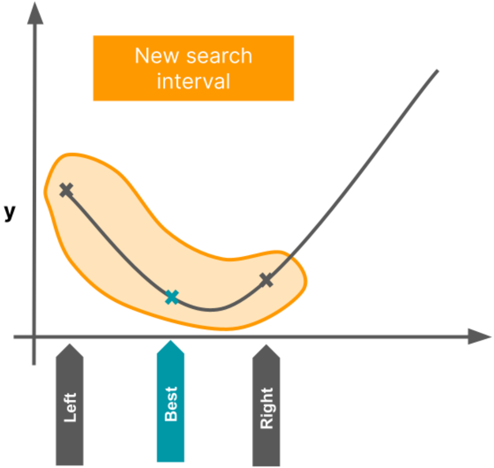

New proposal
y
Left
Best
New
Right

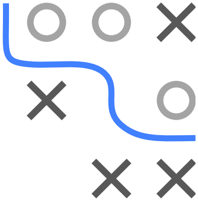


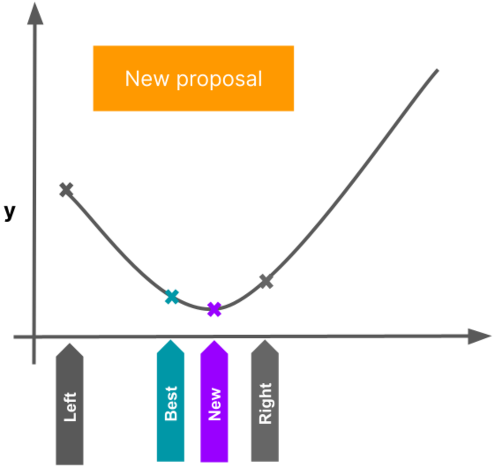

New better than current best?
y
Left
Best
New
Right

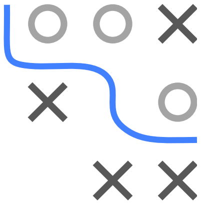

New better than current best?
No...
y
Left
Best
New
Right



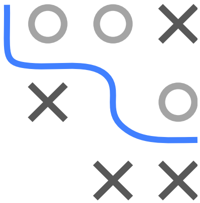

New search
interval
y
Left
Best
Right



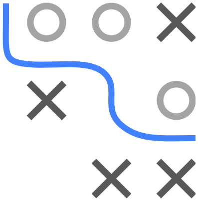

New proposal
y
Left
Best
New
Right

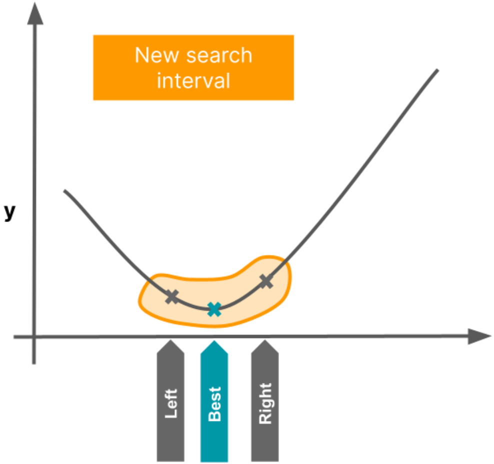


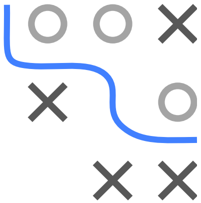

New better than current best?
y
Left
Best
New
Right




New better than current best?
Yes!!!
y
Left
Old best
Best
Right

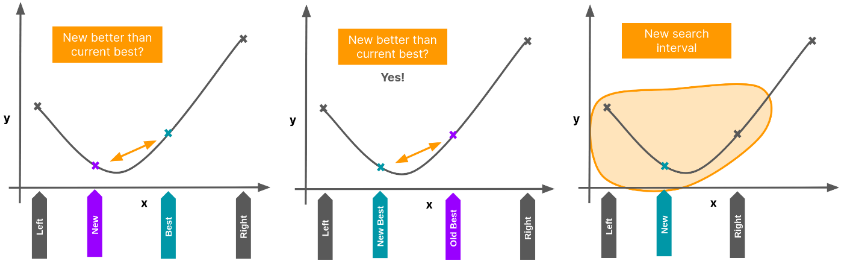



New search interval
y
Left
Best
Right



SIMPLE NESTING PROCEDURE

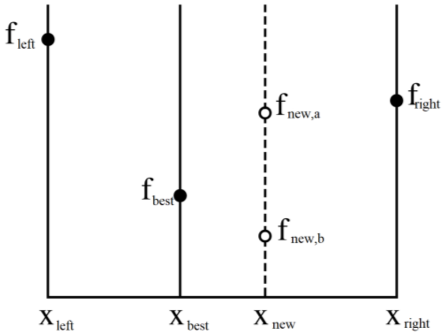
- Initialization: Search interval $(x_{\text{left}}, x_{\text{right}})$, $x_{\text{left}} < x_{\text{right}}$
- Choose x_{best} randomly
- For $t = 0, 1, 2, \ldots$
  - Choose x_{new} randomly in $[x_{\text{left}}, x_{\text{right}}]$
  - If $f(x_{\text{new}}) < f(x_{\text{best}})$:
    - $x_{\text{best}} \leftarrow x_{\text{new}}$
  - New interval: points around x_{best}



GOLDEN RATIO

Key question: How can x_{new} be chosen better than randomly?

- Insight 1: Always in bigger subinterval to maximize reduction
- Insight 2: x_{new} symmetrically to x_{best} for uniform reduction

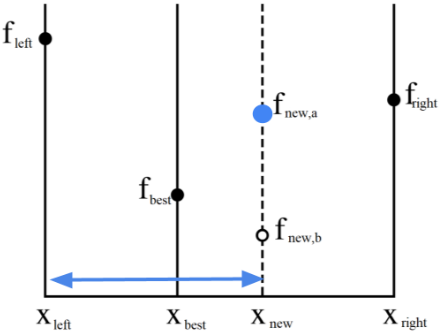


Consider two hypothetical outcomes for x_{new}: $f_{new,a}$ and $f_{new,b}$

GOLDEN RATIO

If outcome is $f_{new,a}$, x_{best} remains best and we search around x_{best}:

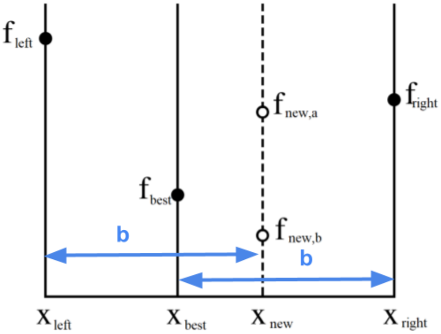
$$[x_{\text{left}}, x_{\text{new}}]$$



GOLDEN RATIO

If $f_{new,b}$ is outcome, x_{new} becomes best point and search around x_{new}:

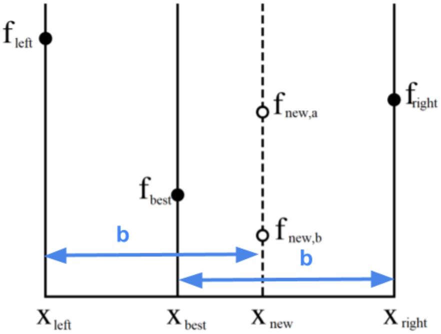
$$[x_{\text{best}}, x_{\text{right}}]$$

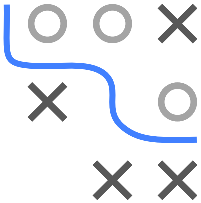


GOLDEN RATIO

For uniform reduction the two potential intervals must be equal-sized:

$$b := x_{\text{right}} - x_{\text{best}} = x_{\text{new}} - x_{\text{left}}$$

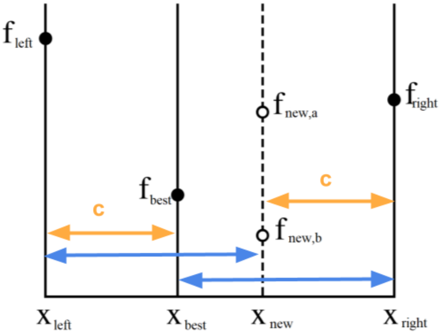


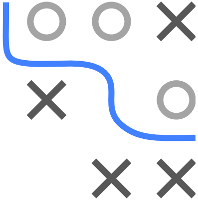


GOLDEN RATIO

One iteration ahead: require again the intervals to be of same size.

$$c := x_{\text{best}} - x_{\text{left}} = x_{\text{right}} - x_{\text{new}}$$

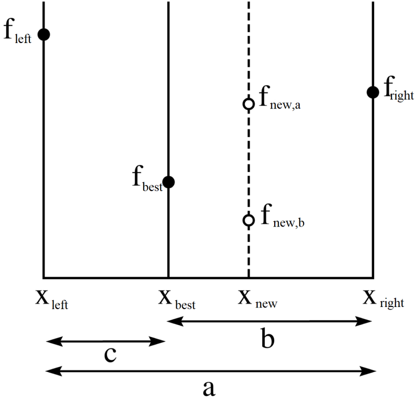




GOLDEN RATIO

To summarize, we require:

$$
\begin{aligned}
a &= x_{\text{right}} - x_{\text{left}}, \\
b &= x_{\text{right}} - x_{\text{best}} = x_{\text{new}} - x_{\text{left}} \\
c &= x_{\text{best}} - x_{\text{left}} = x_{\text{right}} - x_{\text{new}}
\end{aligned}
$$



GOLDEN RATIO

- We require same percentage improvement in each iteration
- For φ reduction factor of interval sizes (a to b, and b to c)

$$\varphi := \frac{b}{a} = \frac{c}{b}$$

$$\varphi^2 = \frac{b}{a} \cdot \frac{c}{b} = \frac{c}{a}$$

- Divide $a = b + c$ by a:

$$\begin{aligned} \frac{a}{a} &= \frac{b}{a} + \frac{c}{a} \\ 1 &= \varphi + \varphi^2 \\ 0 &= \varphi^2 + \varphi - 1 \end{aligned}$$

- Unique positive solution is $\varphi = \frac{\sqrt{5}-1}{2} \approx 0.618$

GOLDEN RATIO

- With x_{new} we always go φ percentage points into the interval
- Given x_{left} and x_{right} it follows

$$\begin{aligned} x_{\text{best}} &= x_{\text{right}} - \varphi(x_{\text{right}} - x_{\text{left}}) \\ &= x_{\text{left}} + (1 - \varphi)(x_{\text{right}} - x_{\text{left}}) \end{aligned}$$

and due to symmetry

$$\begin{aligned} x_{\text{new}} &= x_{\text{left}} + \varphi(x_{\text{right}} - x_{\text{left}}) \\ &= x_{\text{right}} - (1 - \varphi)(x_{\text{right}} - x_{\text{left}}). \end{aligned}$$

GOLDEN RATIO

- Some termination criterion has to be chosen
- A reasonable choice is the absolute error, i.e. the width of the last interval:

$$|x_{\text{best}} - x_{\text{new}}| < \tau$$

- In practice, more complicated termination criteria are usually applied, e.g. *Numerical Recipes in C* ▸ Teukolsky et al. 2017 proposes

$$|x_{\text{right}} - x_{\text{left}}| \leq \tau(|x_{\text{best}}| + |x_{\text{new}}|)$$

as a termination criterion