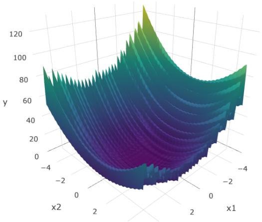
Optimization in Machine Learning

Multi-Start Optimization



Learning goals

- Multimodal functions
- Basins of Attractions
- Simple multi-start procedure

MOTIVATION

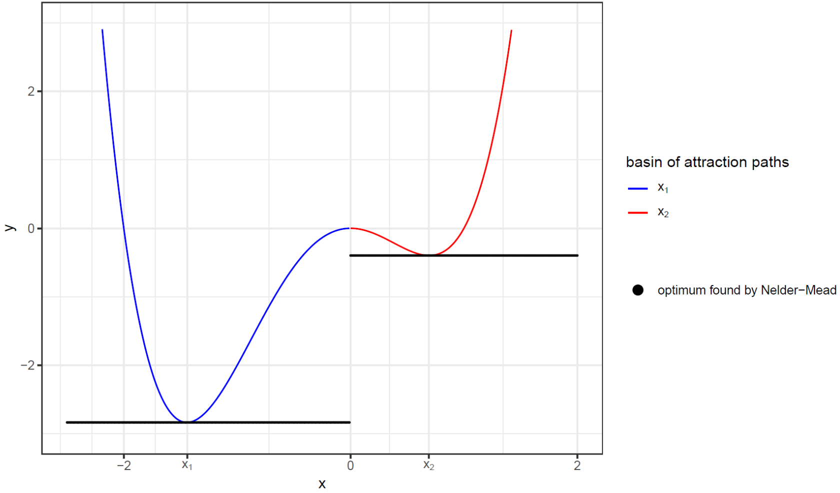
- So far: optimization method for unimodal objective function (exception: simulated annealing)
- If the objective function is actually multimodal, our procedures converge to **local minima**.
- Found optima may be different for different starting values $\boldsymbol{x}^{[0]}$

$\Rightarrow$ **"attraction areas"**

Let $f_1^*, \ldots, f_k^*$ be local minimum values of $f(\cdot)$ with $f_i^* \neq f_j^* \quad \forall i \neq j$. Then the set $\mathcal{A}$ is called basin of attraction of f_i^* for algorithm A.

$$\mathcal{A}(f_i^*, A) = \{\boldsymbol{x} \in S \subseteq \mathbb{R} : f(A(\boldsymbol{x})) \to f_i^*\}$$

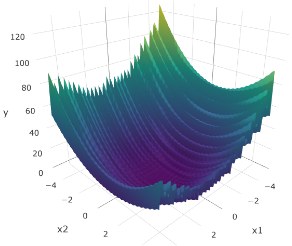
ATTRACTION AREAS



MULTI-STARTS

$$f(\boldsymbol{x}) = \sin^2(3\pi x_1) + (x_1 - 1)^2[1 + \sin^2(3\pi x_2)] + (x_2 - 1)^2[1 + \sin^2(2\pi x_2)]$$

- Global minimum: $f(\boldsymbol{x}^*) = 0$ at $\boldsymbol{x}^* = (1, 1)$
- We optimize the Levy function using the BFGS method with a random starting value of $-2 \leq x_1 \leq 2, -2 \leq x_2 \leq 2$ and note the objective function value of the result after optimization
- We repeat this 100 times



Distribution of the 100 optimization results (y values):

```
## Min.    1st Qu.    Median    Mean    3rd Qu.    Max.
## 0.0000  0.1099     0.5356    2.4351  1.9809     18.3663
```

MULTI-STARTS

Idea: use multiple starting points $\boldsymbol{x}^{[1]}, \dots, \boldsymbol{x}^{[k]}$ for algorithm A

Algorithm Multistart optimization

```
Given: optimization algorithm A(·), f : S ↦ ℝ, x ↦ f(x)
k = 0
repeat
    Draw starting point x^[k] from S (e.g. uniform if S is of finite volume)
    if k = 0 then x̂ = x_0
    end if
    Initialize algorithm with start value x^[k] ⇒ x̃ = A(x^[k])
    if f(x̃) < f(x̂) then x̂ = x̃
    end if
    k = k + 1
until Stop criterion fulfilled
return x̂
```

MULTI-STARTS

BFGS with Multistart gives us the true minimum of the Levy function:

```
iters = 20 # number of starts
xbest = c(runif(1, -2, 2), runif(1, -2, 2))

for (i in 1:iters) {
x1 = runif(1, -2, 2)
x2 = runif(1, -2, 2)
res = optim(par = c(x1, x2), fn = f, method = "BFGS")
}

if (res$value < f(xbest)) {
xbest = res$par
}

xbest
## [1] 1 1
```