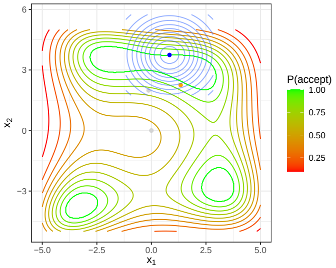
Optimization in Machine Learning

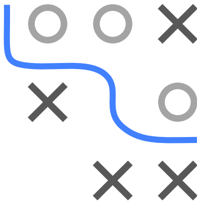
Simulated Annealing



Learning goals

- Motivation
- Metropolis algorithm
- Simulated Annealing

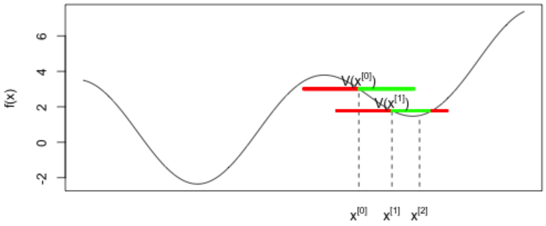
INTRODUCTION



- Heuristics for optimization of complex objectives (multivariate, non-linear, non-convex)
- Procedure for finding good solutions to complex problems
- Does not guarantee optimal/best result (global optimum), but usually good solutions
- Goal for complex optimization problems: avoid “getting stuck” in local optima
- Often used for difficult discrete problems as well
- Local search strategy with random option to accept worse values

SIMPLE STOCHASTIC LOCAL SEARCH

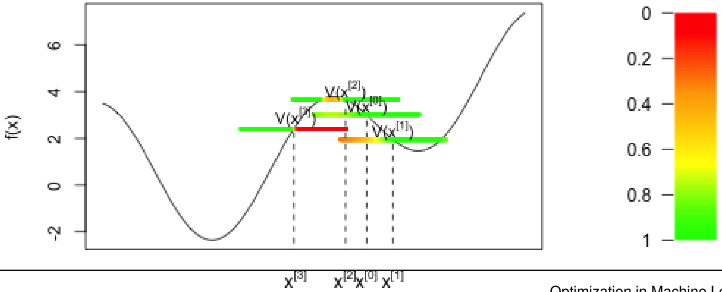
- Given is a multivariate objective function $f(\boldsymbol{x})$
- Define a local neighborhood area $V(\boldsymbol{x})$ for a given $\boldsymbol{x}$
- Sample proposal $\boldsymbol{x}^{[t+1]}$ uniformly at random from $V(\boldsymbol{x}^{[t]})$
- Calculate $f(\boldsymbol{x}^{[t+1]})$
- If $\Delta f = f(\boldsymbol{x}^{[t+1]}) - f(\boldsymbol{x}^{[t]}) < 0$, $\boldsymbol{x}^{[t+1]}$ is accepted as new solution, otherwise a new proposal from neighborhood is sampled



stoch. local search: acceptance range in green and rejection in red

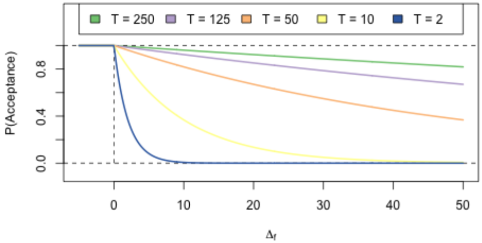
METROPOLIS ALGORITHM

- Simple stochastic local search strongly depends on $\boldsymbol{x}^{[0]}$ and the neighborhood
  $\Rightarrow$ Danger of ending up in local minima
- **Idea:** allow worse candidates with some probability
- **Metropolis:** accept candidates from previous rejection range ($\Delta f > 0$) with probability $\mathbb{P}(\text{accept} \mid \Delta f) = \exp(-\Delta f / T)$
- T denotes “temperature”

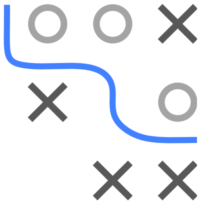


METOPOLIS ALGORITHM

- Parameter T describes temperature/progress of the system
- High temperatures correspond to high probability of accepting worse $\boldsymbol{x}$
- Local minima can be escaped, but no convergence can be achieved at *constant* temperature
- We come across an important principle of optimization: **exploration (high T) vs. exploitation (low T)**



SIMULATED ANNEALING



- Start with high temperature to explore whole space
- Slowly reduce temperature to converge
  $\Rightarrow$ Sequence of descending temperatures $T^{[t]}, t \in \mathbb{N}$
- Procedure is called simulated annealing
- Temperature is often kept constant several iterations in a row to explore the space, then multiplied by coefficient $0 < c < 1$:

$$T^{[t+1]} = c \cdot T^{[t]}$$

- Other strategies possible, for example:

$$T^{[t]} = T^{[0]} \left(1 - \frac{t}{t_{\text{max}}}\right)$$

- Many different strategies for choosing neighborhood
  Strongly depends on objective function

ANALOGY TO METALLURGY

- Simulated annealing draws analogy between a cooling process (e.g. a metal or liquid) and an optimization problem.
- If cooling of a liquid material (amount of atoms) is too fast, it solidifies in suboptimal configuration, slow cooling produces crystals with optimal structure (minimum energy stage)
- Consider atoms of the liquid as a system with many degrees of freedom, analogy to optimization problem of a multivariate function
- Minimum energy stage corresponds to optimum of objective function