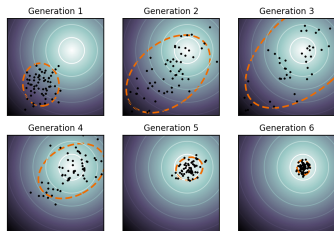
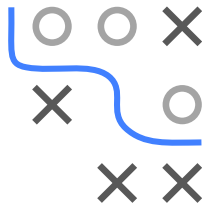


Optimization in Machine Learning

Evolutionary Algorithms

CMA-ES Algorithm

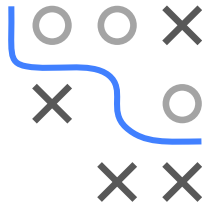


Learning goals

- CMA-ES strategy
- Estimation of distribution
- Step size control

CMA-ES: BASIC METHOD - ITERATION 1

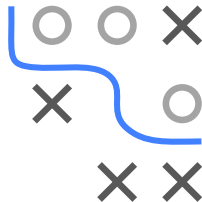
- 1 Initialize $\mathbf{m}^{[0]}$, $\sigma^{[0]}$ problem-dependent and $\mathbf{C}^{[0]} = \mathbf{I}_d$



CMA-ES: BASIC METHOD - ITERATION 1

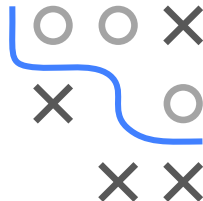
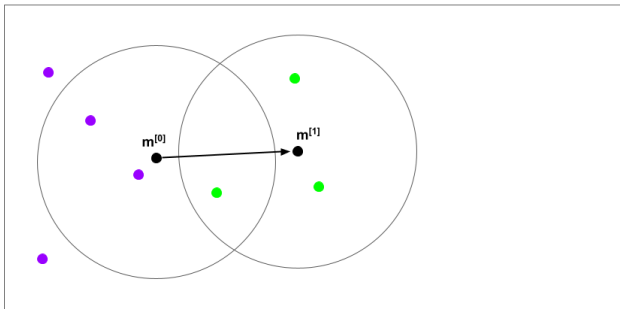
- ❶ **Sample** λ offspring from distribution

$$\mathbf{x}^{[1](i)} = \mathbf{m}^{[0]} + \sigma^{[0]} \mathcal{N}(\mathbf{0}, \mathbf{C}^{[0]})$$



CMA-ES: BASIC METHOD - ITERATION 1

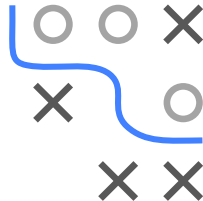
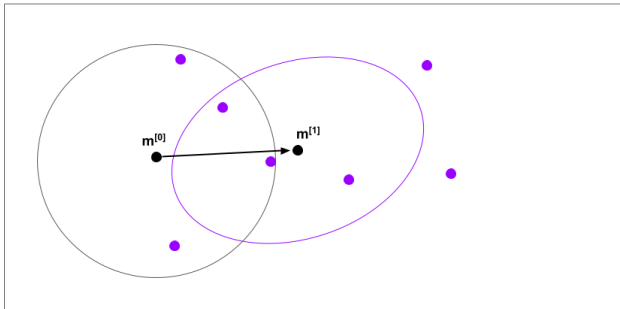
3 Update mean



Movement towards the new distribution with mean $\mathbf{m}^{[1]} = \mathbf{m}^{[0]} + \sigma^{[0]} \mathbf{y}_w^{[1]}$.

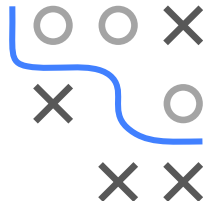
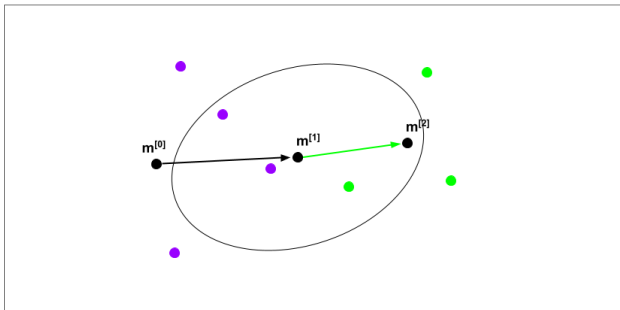
CMA-ES: BASIC METHOD - ITERATION 2

- 1 **Sample** from distribution for new generation



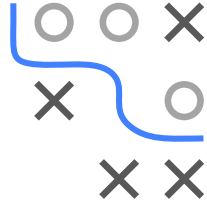
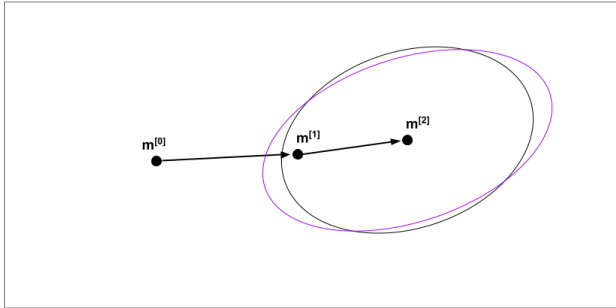
CMA-ES: BASIC METHOD - ITERATION 2

- ② Selection and recombination of $\mu < \lambda$ best-performing offspring
- ③ Update mean



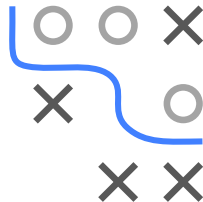
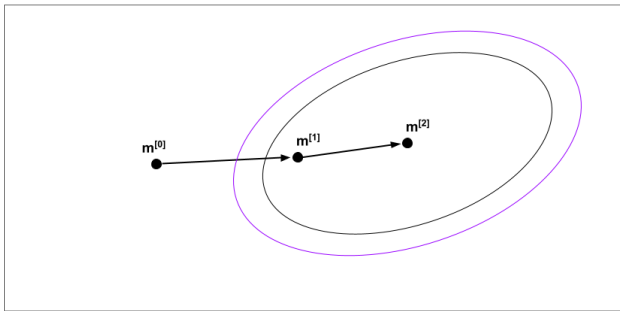
CMA-ES: BASIC METHOD - ITERATION 2

2 Update covariance matrix



CMA-ES: BASIC METHOD - ITERATION 2

- ② **Update step-size** exploiting correlation in history of steps.
steps point in similar direction \Rightarrow increase step-size
 \Rightarrow decrease step-size



UPDATING C: FULL UPDATE

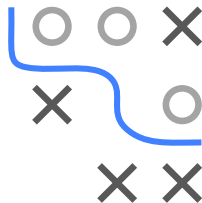
Full CMA update of \mathbf{C} combines rank- μ update with a rank-1 update using exponentially smoothed evolution path $\mathbf{p}_c \in \mathbb{R}^d$ of successive steps and learning rate c_1 :

$$\mathbf{p}_c^{[0]} = \mathbf{0}, \quad \mathbf{p}_c^{[t+1]} = (1 - c_1)\mathbf{p}_c^{[t]} + \sqrt{\frac{c_1(2 - c_1)}{\sum_{i=1}^{\mu} w_i^2}} \mathbf{y}_w$$

Final update of \mathbf{C} is

$$\mathbf{C}^{[t+1]} = (1 - c_1 - c_{\mu} \sum_j w_j) \mathbf{C}^{[t]} + c_1 \underbrace{\mathbf{p}_c^{[t+1]} (\mathbf{p}_c^{[t+1]})^T}_{\text{rank-1}} + c_{\mu} \underbrace{\sum_{i=1}^{\mu} w_i \mathbf{y}_{i:\lambda}^{[t+1]} (\mathbf{y}_{i:\lambda}^{[t+1]})^T}_{\text{rank-}\mu}$$

- Correlation between generations used in rank-1 update
- Information from entire population is used in rank- μ update



UPDATING σ : METHODS STEP-SIZE CONTROL

- **1/5-th success rule**: increases the step-size if more than 20 % of the new solutions are successful, decrease otherwise
- **σ -self-adaptation**: mutation is applied to the step-size and the better - according to the objective function value - is selected
- **Path length control via cumulative step-size adaptation (CSA)**

Intuition:

- Short cumulative step-size \triangleq steps cancel \rightarrow decrease $\sigma^{[t+1]}$
- Long cumulative step-size \triangleq corr. steps \rightarrow increase $\sigma^{[t+1]}$

