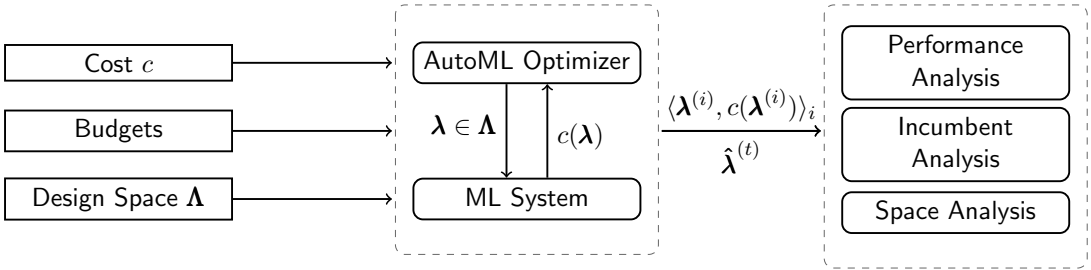
AutoML: Interpretability

Studying the AutoML Optimization Process

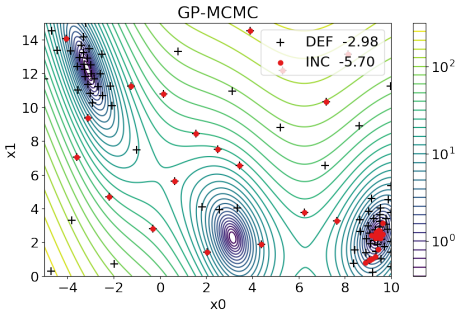
Bernd Bischl Frank Hutter Lars Kotthoff
Marius Lindauer Joaquin Vanschoren

Idea



$\rightsquigarrow$ focus on how the AutoML optimizer samples from the design space $\mathbf{\Lambda}$

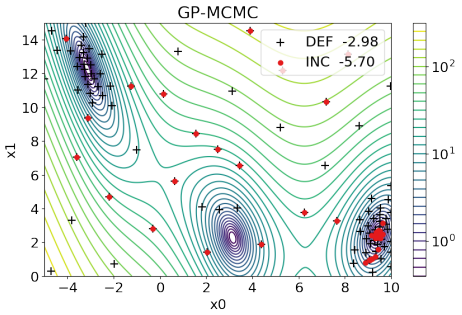
Analyzing the Sampling Behavior



Source: [Lindauer et al. 2019]

- Plot of a 1D or 2D function

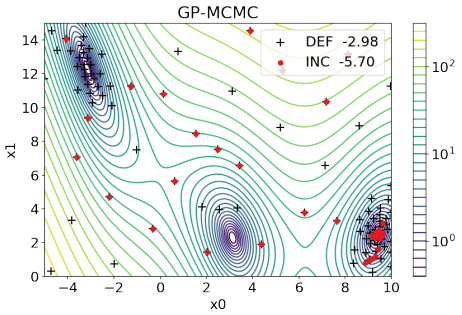
Analyzing the Sampling Behavior



Source: [Lindauer et al. 2019]

- Plot of a 1D or 2D function
- Background shows the ground truth (real function values)

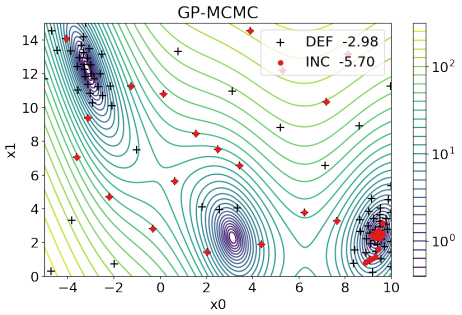
Analyzing the Sampling Behavior



Source: [Lindauer et al. 2019]

- Plot of a 1D or 2D function
- Background shows the ground truth (real function values)
- Dots are sampled points in the search space

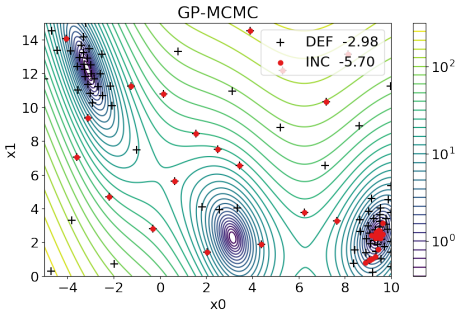
Analyzing the Sampling Behavior



Source: [Lindauer et al. 2019]

- Plot of a 1D or 2D function
- Background shows the ground truth (real function values)
- Dots are sampled points in the search space
- Typical approach in Bayesian Optimization community

Analyzing the Sampling Behavior

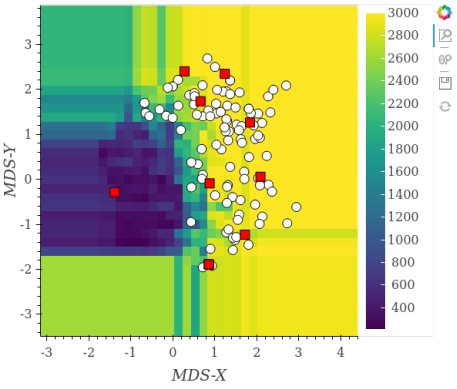


Source: [Lindauer et al. 2019]

- Plot of a 1D or 2D function
- Background shows the ground truth (real function values)
- Dots are sampled points in the search space
- Typical approach in Bayesian Optimization community

$\rightsquigarrow$ Impossible for higher dimensional problems?

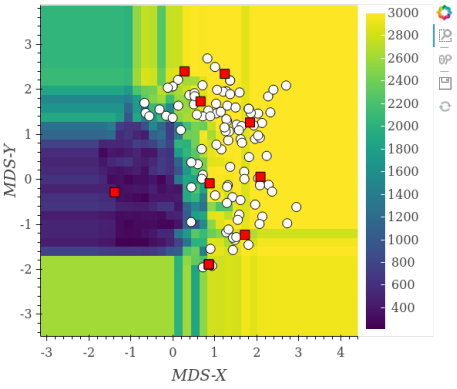
Analyzing the Sampling Behavior in N-D



Source: [Lindauer et al. 2019]

- Same idea as before
  but we have to project N-D into 2-D

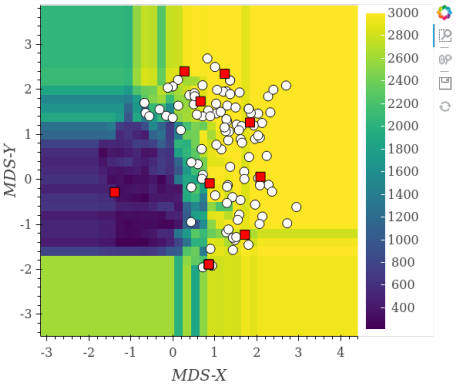
Analyzing the Sampling Behavior in N-D



Source: [Lindauer et al. 2019]

- Same idea as before
  but we have to project N-D into 2-D
  1. Use an MDS to project down to 2-D

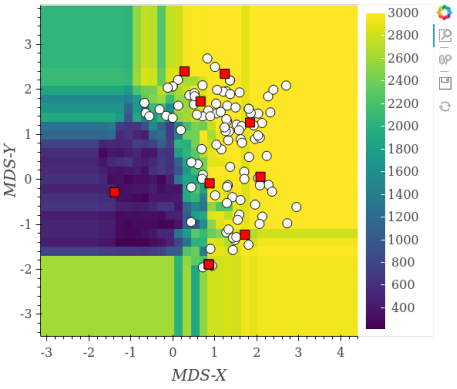
Analyzing the Sampling Behavior in N-D



Source: [Lindauer et al. 2019]

- Same idea as before
  but we have to project N-D into 2-D
  1. Use an MDS to project down to 2-D
  2. Each dot is single hyperparameter configuration

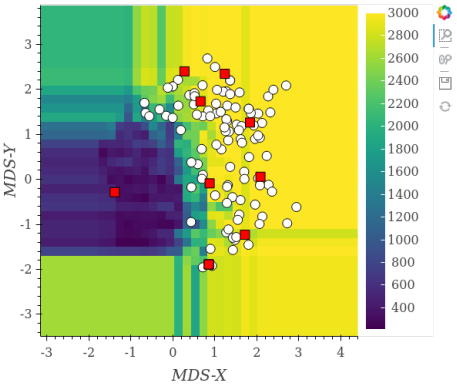
Analyzing the Sampling Behavior in N-D



Source: [Lindauer et al. 2019]

- Same idea as before
  but we have to project N-D into 2-D
  1. Use an MDS to project down to 2-D
  2. Each dot is single hyperparameter configuration
  3. Red squares are intermediate incumbents

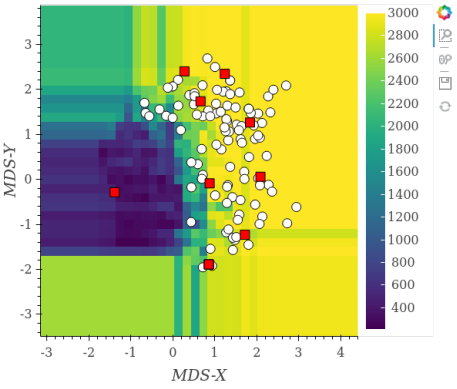
Analyzing the Sampling Behavior in N-D



Source: [Lindauer et al. 2019]

- Same idea as before
  but we have to project N-D into 2-D
  1. Use an MDS to project down to 2-D
  2. Each dot is single hyperparameter configuration
  3. Red squares are intermediate incumbents
  4. The background is colored wrt a performance-estimate
     (e.g., reusing model fitted during BO)

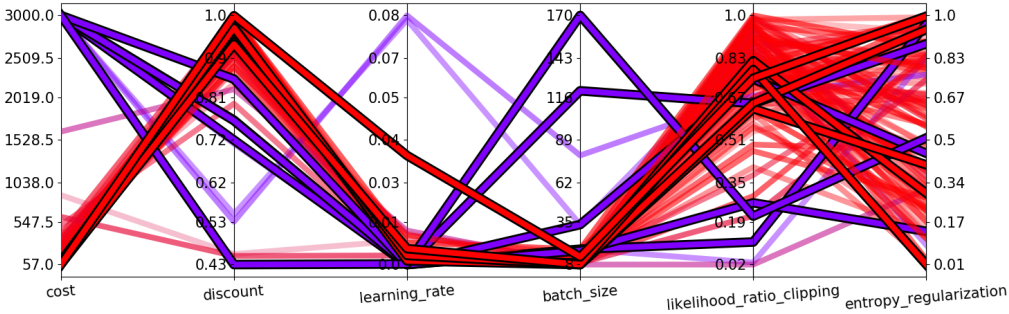
Analyzing the Sampling Behavior in N-D



Source: [Lindauer et al. 2019]

- Same idea as before
  but we have to project N-D into 2-D
  1. Use an MDS to project down to 2-D
  2. Each dot is single hyperparameter configuration
  3. Red squares are intermediate incumbents
  4. The background is colored wrt a performance-estimate (e.g., reusing model fitted during BO)
  5. Extension: Animation by showing how points get added over time

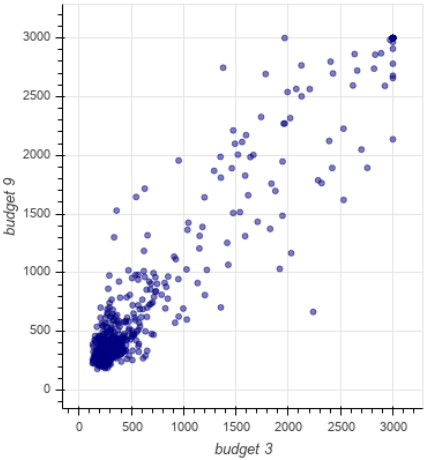
Parallel Coordinate Plot [Golovin et al 2017]



Source: [Lindauer et al. 2019]

- Each coordinate is one hyperparameter;
- Except the most left one: cost or loss

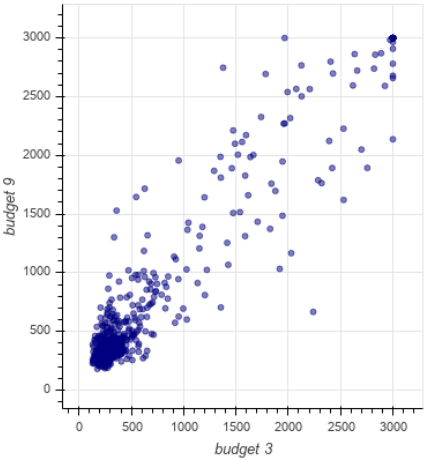
Multi-Fidelity Checks



Source: [Lindauer et al. 2019]

- Challenge of multi-fidelity approaches:
  - How to choose the fidelities (a.k.a. budgets)

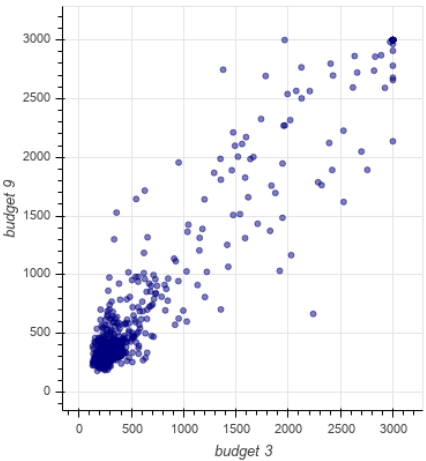
Multi-Fidelity Checks



Source: [Lindauer et al. 2019]

- Challenge of multi-fidelity approaches:
  - How to choose the fidelities (a.k.a. budgets)
- Important Property:
  - Decisions on small budgets should be reasonable for higher budgets

Multi-Fidelity Checks



Source: [Lindauer et al. 2019]

- Challenge of multi-fidelity approaches:
  - How to choose the fidelities (a.k.a. budgets)
- Important Property:
  - Decisions on small budgets should be reasonable for higher budgets
- Analysis:
  1. Scatter plot of performance on Budget X vs. Budget Y
  2. Each dot is sampled hyperparameter configuration
  3. Compute rank correlation (here: 0.69)