RealML @ ICML 2020

Cost-Aware Bayesian Optimization via Information Directed Sampling

Biswajit Paria, Willie Neiswanger, Ramina Ghods,
Jeff Schneider, Barnabás Póczos

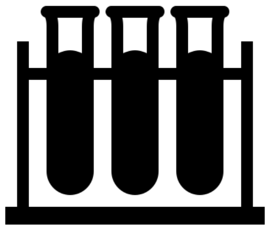
Carnegie Mellon University
School of Computer Science

Cost Aware Bayesian Optimization

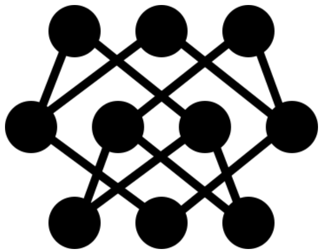
Given: Blackbox function $f(x)$ taking cost $\lambda(x)$ to evaluate.

Goal: Optimize $f(x)$ while minimizing the total cost of all evaluations.

Examples:



Cost of chemicals in a chemistry experiment.



Training time of a neural network.

Related Setting: Multi-Fidelity BO

Our Approach

Existing approaches include BOCA, Multi-Fidelity MES, CostEI

Main advantages of our approach:

- Does not require expensive entropy computations.
    - ❖ Works with a variety of Bayesian models beyond GPs. Example: Bayesian neural networks.
- Theoretically sound approach. Enjoys optimal regret bounds.
- Conceptually simple.

Information Directed Sampling (IDS)

- Principled way to balance exploration and exploitation.
- Balances *expected regret* and *information gain.*

$$\min_{\pi} \frac{\mathbb{E}_{x\sim\pi}\left[f(x^*) - f(x) \mid D_t\right]^2}{\mathbb{E}_{x\sim\pi}\mathrm{IG}[x^*, x \mid D_t]}$$

Expected Regret

Information Gain

CostIDS

- Balances *expected regret, information gain,* and experimental cost.

$$\min_{x} \lambda(x)\left(\frac{\mathbb{E}[f(x^*) \mid D_t] - \mu_t(x)}{\sigma_t(x)}\right)^2$$

Expected Regret

Information Gain

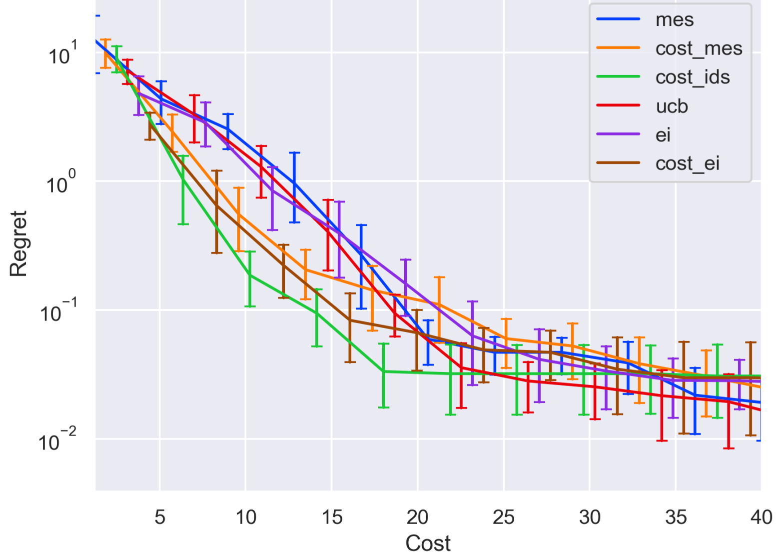
Cost

$\mu_t(x)$: Posterior mean conditioned on D_t

$\sigma_t(x)$: Posterior variance conditioned on D_t

Results

- Promising empirical results on a synthetic function.



Objective function:
Modified Branin function

Further details in the paper