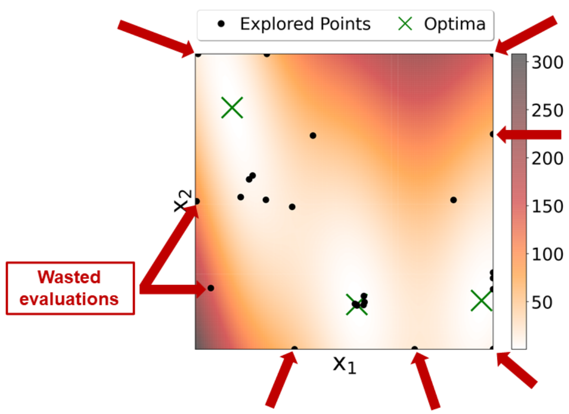
Prior-guided Bayesian Optimization

Artur Souza, Luigi Nardi, Leonardo B. Oliveira, Kunle Olukotun, Marius Lindauer, Frank Hutter

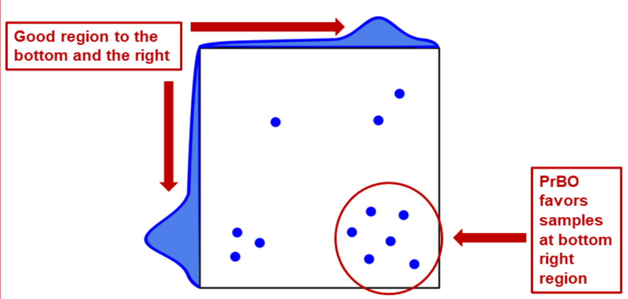
Motivation

While Bayesian Optimization (BO) is a very popular method for optimizing expensive black-box functions, it fails to leverage the experience of domain experts. This causes BO to waste function evaluations on design choices that the expert already knows to work poorly.



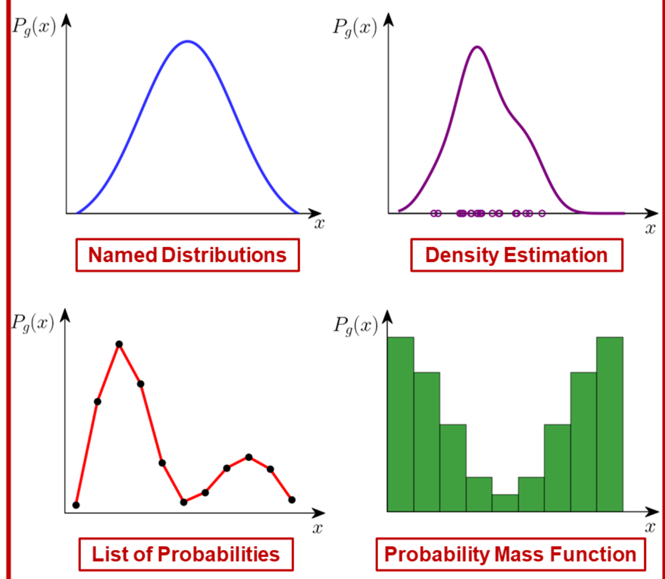
To address this issue, we introduce Prior-guided Bayesian Optimization (PrBO), a novel BO variant that combines user prior knowledge with a probabilistic model of the observations made. Additionally:

- PrBO bridges the TPE methodology and standard BO probabilistic models.
- PrBO is flexible w.r.t. how the prior is defined, allowing previously hard-to-inject priors.
- PrBO gives more importance to the model as iterations progress, gradually forgetting the prior.



Priors

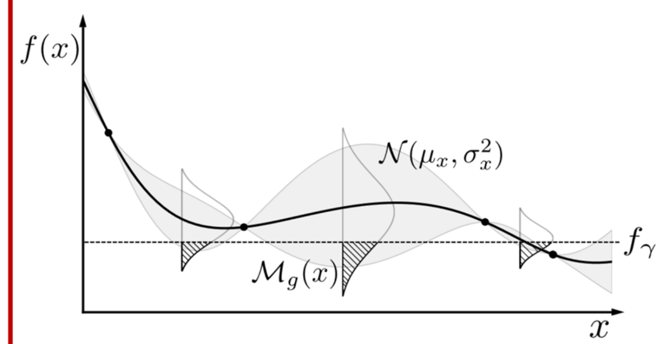
Priors take the form of a prior distribution on where in the input space users expect to find "good" function values. PrBO is very flexible w.r.t. how these priors are defined. E.g. priors can be defined via:



Model

PrBO's model quantifies the probability of a point being good, according to a standard probabilistic model (e.g. a GP or RF). As in Tree-structured Parzen Estimators [1], we define configurations as "good" if their observed value is below a certain quantile γ of the observed function values. We then use Probability of Improvement [2] to compute the probability of the function value lying below this quantile:

$$\mathcal{M}_g(\boldsymbol{x}) = p(f(\boldsymbol{x}) < f_\gamma | \boldsymbol{x}) = \Phi\left(\frac{f_\gamma - \mu_{\boldsymbol{x}}}{\sigma_{\boldsymbol{x}}}\right),$$

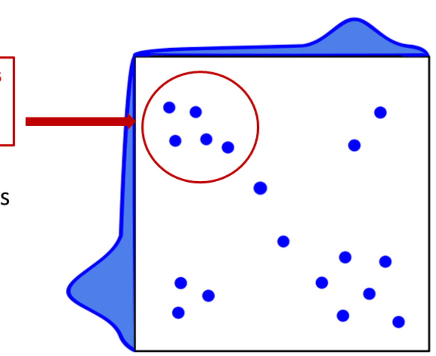


Pseudo-posterior

PrBO combines prior and model into a pseudo-posterior on "good" points. This pseudo-posterior represents the updated beliefs on where we can find "good" points. The pseudo-posterior is computed as the product of the prior and the model:

$$g(\boldsymbol{x}) \propto P_g(\boldsymbol{x}) \mathcal{M}_g(\boldsymbol{x})^{\frac{t}{\beta}}$$

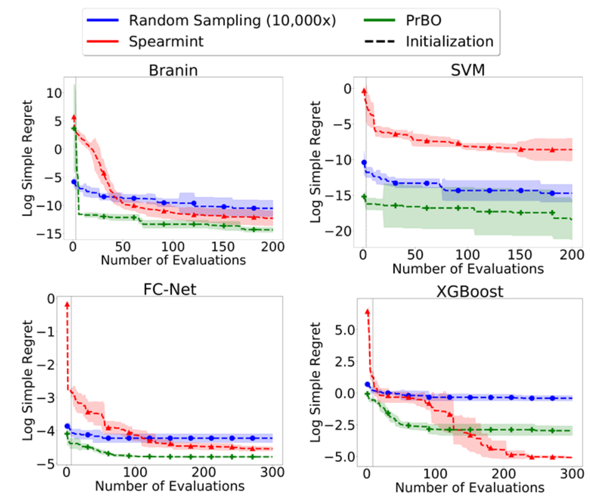
Model weight ensures that PrBO forgets misleading priors



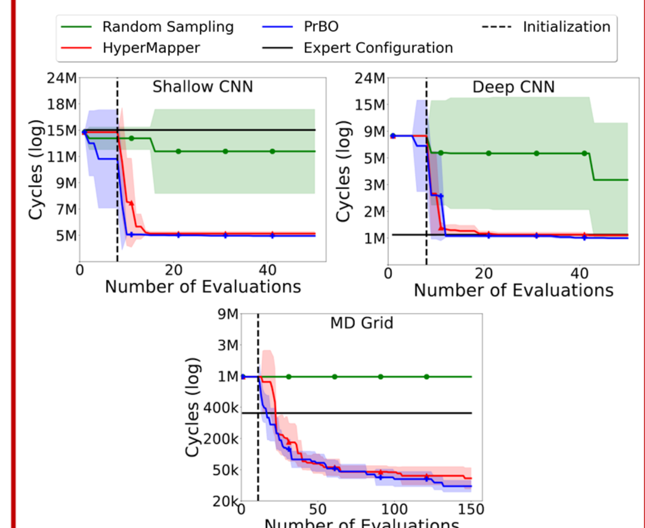
t/β controls how much weight is given to the model versus the prior. This ensures that PrBO puts more emphasis on the model as it observes more data and becomes more accurate and allows PrBO to recover from misleading priors. Similar to, and inspired by Bayesian models, the data ultimately washes out the prior.

Results

Synthetic benchmarks:



Real-world hardware design application:



References

[1] Bergstra *et al. Algorithms for Hyper-parameter optimization.* In NeurIPS, 2011.
[2] H. J. Kushner. *A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise.* Journal of Basic Engineering, 86(1):97-106, 1964.