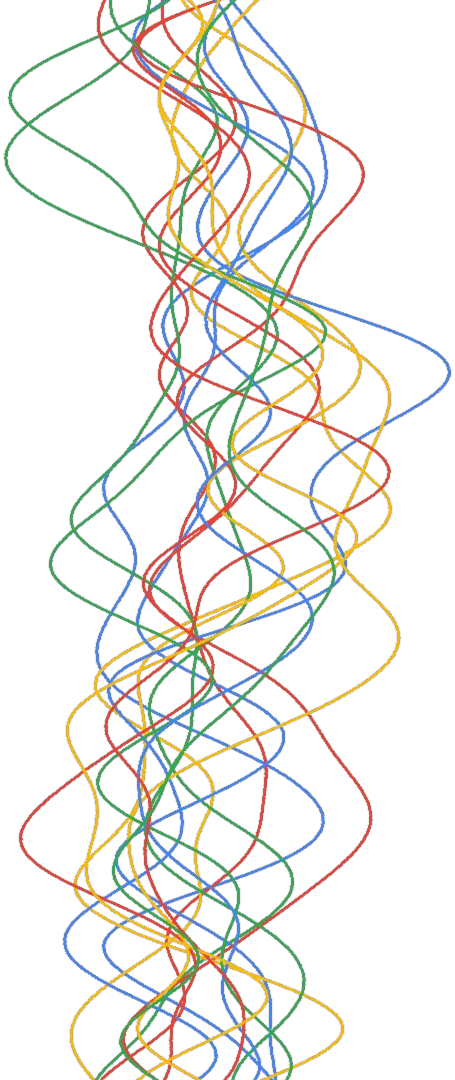


Pre-trained Gaussian processes for Bayesian optimization

Journal of Machine Learning Research (JMLR), 2024

Zi Wang

Research Scientist

https://ziw.mit.edu/

Zi Wang

George Dahl

Kevin Swersky

Chansoo Lee

Zack Nado

Justin Gilmer

Jasper Snoek

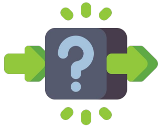
Zoubin Ghahramani

Google DeepMind

Bayesian optimization for global optimization of black-box functions

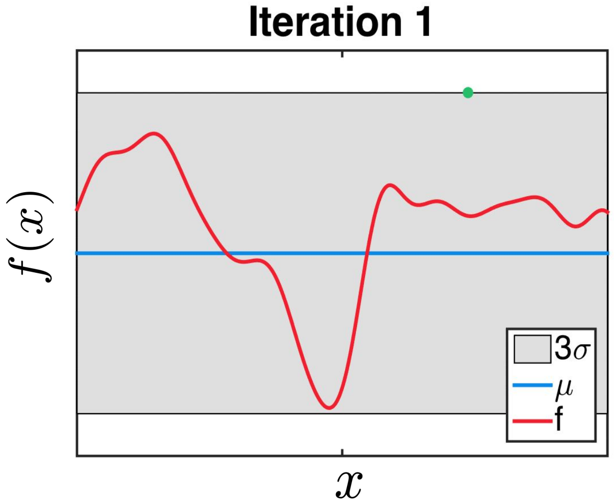
Designing experiments
as a domain expert

Design parameters
x



Measure of performance
$f(x)$

E.g., hyperparameter tuning, protein engineering, synthetic chemistry, robot learning, baking cookies, choosing careers...

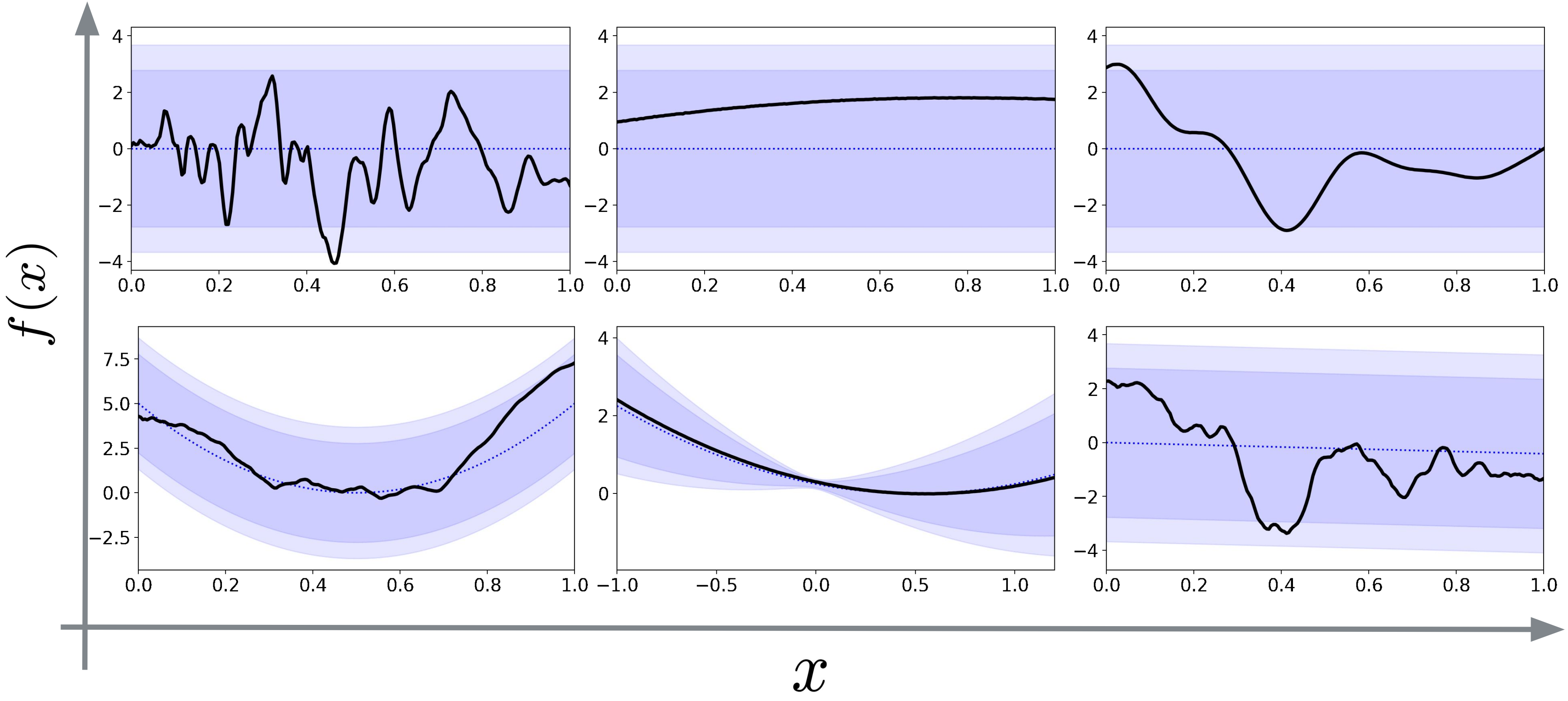


Our problem: Optimize a black-box function.

$$\arg\max_x \; f(x)$$

image: Flaticon.com

Which Gaussian processes to use as the prior? $f \sim \mathcal{GP}(\mu, k)$



Visualizations of interfaces are from https://research.google/blog/pre-trained-gaussian-processes-for-bayesian-optimization/

Challenges in BayesOpt

- BayesOpt is theoretically strong, but its performance can suffer if the GP prior isn't well-suited to the problem.
- Users often need to carefully select GP mean and kernel parameters.

Our interface: HyperBO

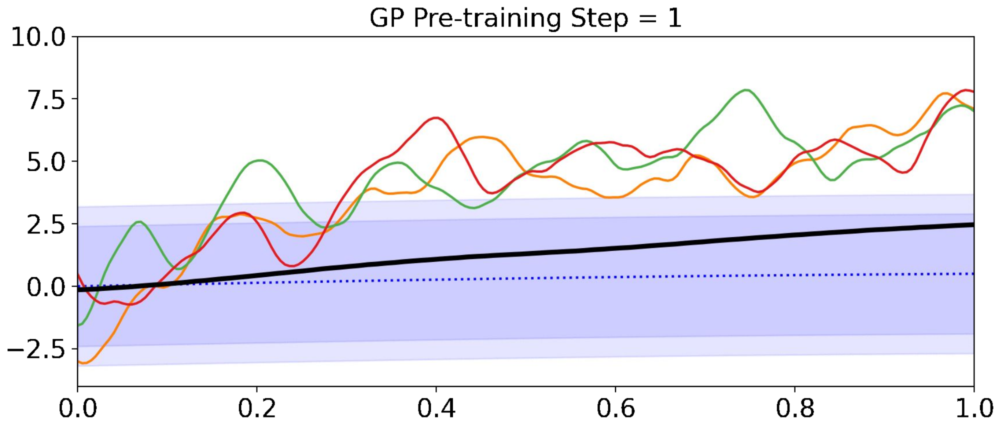
Selection of related tasks for **pre-training a GP**.

- Better alignment with ground truth user belief of the function.*
- Improve the performance of BayesOpt methods.

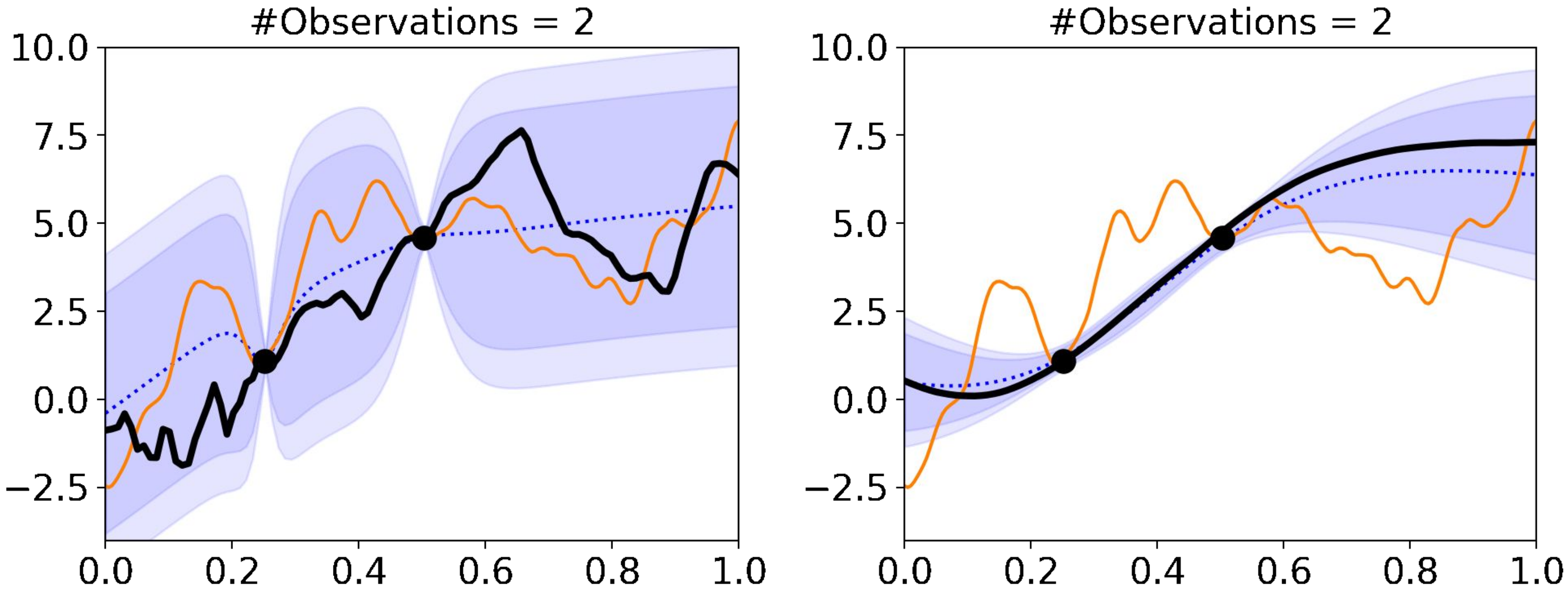
* Under some assumptions.

Model pre-training in function spaces

- Approximations for objective function KL(ground truth GP || model)
  - Empirical KL divergence (EKL): divergence between an empirical estimate of the ground truth model and the pre-trained model.
  - Negative log likelihood (NLL): sum of negative log likelihoods of the pre-trained model for all training functions.



Pre-trained GPs achieve better posterior alignment



GP Pre-training enhances the performance of Bayesian Optimization

Theoretical guarantees (informal)

1. **Bounded posterior:** The pre-trained GP posterior mean and variance are bounded by the ground truth posterior mean and variance.

2. **Near-zero regret bound:** The regret of BayesOpt with a pre-trained GP is bounded.

With probability $1 - \delta$,

Approaches 0

$$R_T < O\left(\sqrt{\frac{T}{N-T-1}} + \sqrt{\log \frac{1}{\delta}}\right) O(\sqrt{\rho_T / T} + \sigma_*)$$

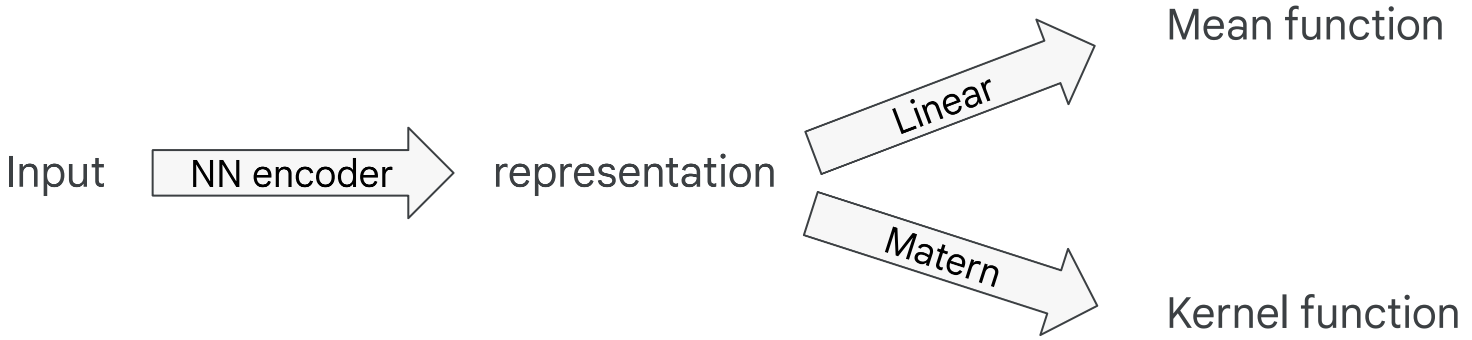
Simple regret

#Training functions

#BO iterations

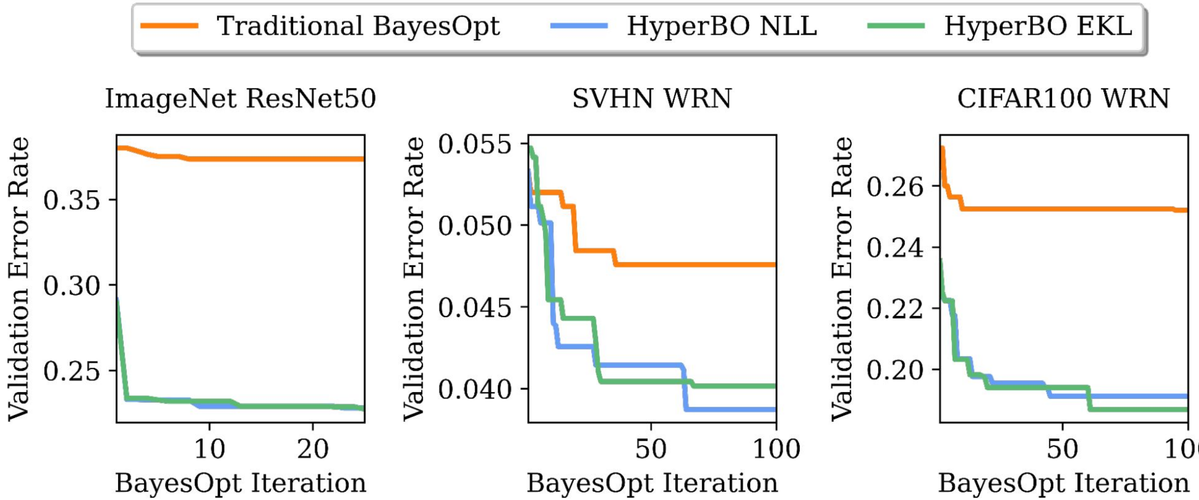
Observation noise

Example setup of HyperBO in our experiments



GP Pre-training enhances the performance of Bayesian Optimization

- PD1 dataset: ~50,000 hparam evaluations of near-SOTA deep learning models on image, text, and protein sequence datasets.
- >3x more efficient than the best competing methods.



https://github.com/google-research/hyperbo/

HyperBO

Gaussian process pre-training makes BayesOpt more effective and easier to use

Contact: wangzi@google.com
Website: https://ziw.mit.edu/

https://github.com/google-research/hyperbo/
https://github.com/google-research/gpax

Google DeepMind