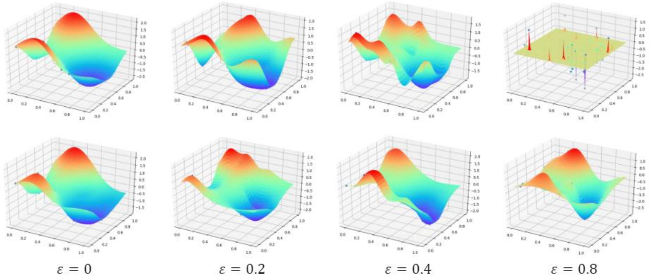
Neighbor Regularized Bayesian Optimization for Hyperparameter Optimization

Lei Cui Yangguang Li Xin Lu Dong An Fenggang Liu

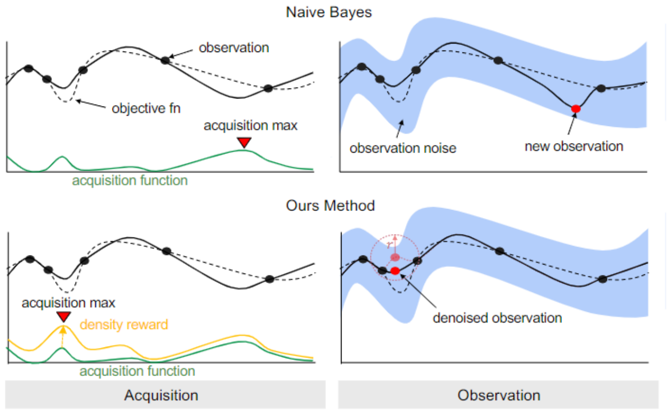
Bayesian Optimization (BO) is a common solution to search optimal hyperparameters based on sample observations of a machine learning model. **Existing BO algorithms could converge slowly even collapse when the potential observation noise misdirects the optimization.**

In this paper, we propose a novel BO hyperparameter optimization algorithm called **Neighbor Regularized Bayesian Optimization (NRBO):**

- We propose a neighbor-based regularization to smooth each sample observation, which could reduce the observation noise efficiently without any extra training cost.
- We further design a density-based acquisition function to adjust the acquisition reward and obtain more stable statistics.
- We design a adjustment mechanism to ensure the framework maintains a reasonable regularization strength and density reward conditioned on remaining computation resources.



Fitting surrogate model on the same dataset with different noise level. From left to right: Gradually increased noise level ε, from 0 to 0.8. Top row: A typical collapse case of Bayesian optimization with Gaussian process. Bottom row: Neighbor regularized Bayesian optimization.



- In observation stage, neighbor regularized mechanism is introduced to smooth the observation noise and release the burden of repetitive observation.
- In acquisition stage, we propose a density-based acquisition function to accelerate the acquisition process, in which adjacent sample points in the neighbor are considered.

Method	MLP_digits	RF_breast	SVM_digits	SVM_wine	ada_breast	ada_digits	linear_breast	*Avg._score*	*Avg._norm*
Random	98.63	92.07	94.95	81.19	93.75	69.26	89.27	90.93	1.07
Hyperopt	100.29	96.50	96.87	88.72	98.96	73.87	92.85	96.53	0.411
Opentuner	100.76	92.07	98.43	92.61	98.96	100.21	94.39	93.16	0.81
Nevergrad	100.11	98.83	100.07	70.30	98.96	76.09	94.55	93.95	0.72
Pysot	101.13	95.34	97.56	99.87	95.83	96.55	97.80	98.34	0.20
Skopt	100.08	98.60	96.86	96.24	93.75	76.79	89.88	96.23	0.45
Turbo	101.10	96.74	98.53	100	95.83	91.45	97.56	98.28	0.20
HEBO	101.38	97.67	101.70	96.24	98.96	112.88	95.82	99.94	0.01
NRBO	101.49	100.00	103.53	103.63	100.00	113.01	101.01	100.34	-0.04

The performance of 9 optimizers on bayesmark benchmark

Method	ImageNet	VOC	CIFAR10	CIFAR100	Stanford Car	COCO
Random	62.00	75.02	95.28	81.96	87.09	31.56
HEBO	62.07(+0.07)	74.88(-0.14)	95.35(+0.07)	81.91(-0.05)	87.56(+0.47)	32.34(+0.78)
NRBO	62.22(+0.22)	77.12(+2.10)	95.43(+0.15)	82.16(+0.20)	87.94(+0.85)	33.92(+1.36)

Experiment results on popular computer vision tasks.

Exp. Index	Task	Random	HEBO	NRBO
1	proxy	67.268	65.808	66.490
	full	68.324	67.660	68.362
2	proxy	65.644	67.258	66.972
	full	67.858	68.642	68.562
3	proxy	66.290	66.908	67.058
	full	67.668	68.494	68.446
4	proxy	66.916	66.116	67.020
	full	68.638	67.848	68.606
Avg.	proxy	66.530	66.523	**66.885**
	full	68.122	68.161	**68.494**

Exp. Index	Task	Random	HEBO	NRBO
1	proxy	32.283	32.070	34.853
	full	34.549	34.438	36.952
2	proxy	32.085	32.334	33.231
	full	33.773	34.609	35.634
3	proxy	31.705	33.571	34.410
	full	34.123	35.997	36.199
4	proxy	30.147	31.395	33.222
	full	32.487	33.585	35.307
Avg.	proxy	31.555	32.343	**33.929**
	full	33.733	34.657	**36.023**

(Left) Proxy task and full-training task results on ImageNet. (Right) Proxy task and full-training task results on COCO.
The hyperparameter searched by NRBO on proxy task still take the lead when transferred to full training tasks.