Hyperparameter Optimization

AutoML Assignment 1

Jiaxuan Yu, Arief Rahman

1 Theory questions

1. Surrogate model

   In this experiment, surrogate model is a predictive model to estimate the machine learning models performance in hyperparameter optimization (HPO). It helps to reduce the computational cost of HPO by not fully training each machine learning model. The external surrogate model is trained on the learning curve database (LCDB) dataset and it is implemented to predict the performance of a given configuration to be used as the ground truth.

2. SMBO implementation

   The Sequential model-based optimization (SMBO) is an instantiation of Bayesian Optimization of which the key design principle is to approximate and evaluate the unknown parts of the objective function with internal surrogate model and use acquisition function to choose the next point(s) for evaluation.[1].

   - Internal surrogate model: the Gaussian process is commonly used for its closed form posterior predictive equations[3].
   - Acquisition function: Expected Improvement(EI) acquisition function is wildly used due to its closed-form expression under a GP surrogate model[3].

     EI is defined as:

     $$EI_t(x) = E[[F(x) - f_t^*]^+] = \sigma(x)\phi(\frac{\mu(x) - f_t^*}{\sigma(x)}) + (\mu(x) - f_t^*)\Phi(\frac{\mu(x) - f_t^*}{\sigma(x)}) \tag{1}$$

     where Φ and ϕ are the CDF and PDF of a standard normal distribution and f_t^* is the current best performance. After the configuration is predicted by Gaussian Process we can get the $\mu(x)$ and $\sigma(x)$ which are the mean and standard deviation of the normal distribution.

3. Successive Halving

   Successive halving is a resource allocation method, which, in each iteration, prunes poorly-performing configurations and progressively allocates more resources to the the best half of the configurations and continue until we only have one configuration with the best performance[2].

   In this experiment, we need to use the anchor functionality of the external surrogate (that can make predictions across different anchors).

2 Empirical questions

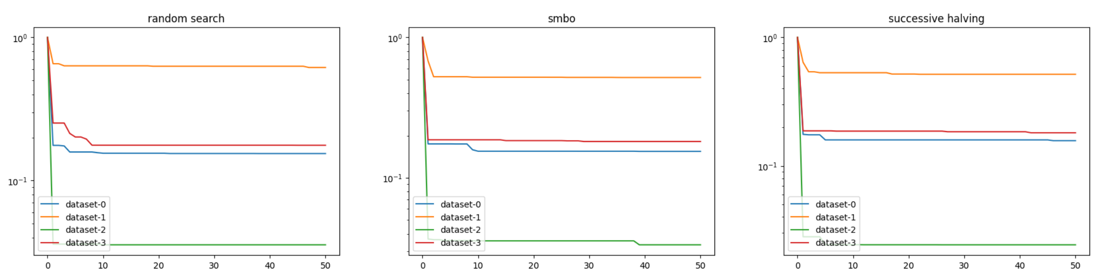
1. The working of the surrogate model

To build external surrogate model, we preprocess the data, imputing numerical features and converting categorical features into one-hot encoded numerical form. A pipeline serves as the core component, combining the preprocessor and RandomForestRegressor. We output mean squared error and R2 score of each independently trained surrogate model for evaluation.

Table 1 shows the working of the surrogate model on a holdout set and its spearman correlation. Surrogate models across all datasets fit well which indicates a strong predictive ability. The config_performances_dataset-6 is fitted the best with the highest R2 score, spearman correlation and MSE value to be 0.

algorithm	dataset	mse	R2 score	spearman correlation	best performance
random search	lcdb_configs	0.0003	0.9925	0.9965	0.1542
	config_performances_dataset-1457	0.0000	0.9969	0.9889	0.5411
	config_performances_dataset-11	0.0000	0.9975	0.9979	0.0268
	config_performances_dataset-6	0.0000	0.9998	0.9998	0.1764
SMBO	lcdb_configs	0.0003	0.9940	0.9957	0.1529
	config_performances_dataset-1457	0.0000	0.9958	0.9911	0.5142
	config_performances_dataset-11	0.0000	0.9986	0.9983	0.0250
	config_performances_dataset-6	0.0000	0.9998	0.9998	0.1764
successive halving	lcdb_configs	0.0004	0.9921	0.9960	0.1541
	config_performances_dataset-1457	0.0001	0.9951	0.9909	0.5150
	config_performances_dataset-11	0.0000	0.9983	0.9984	0.0252
	config_performances_dataset-6	0.0000	0.9998	0.9998	0.1764

Table 1: Performance of surrogate model using three optimization algorithms on different datasets



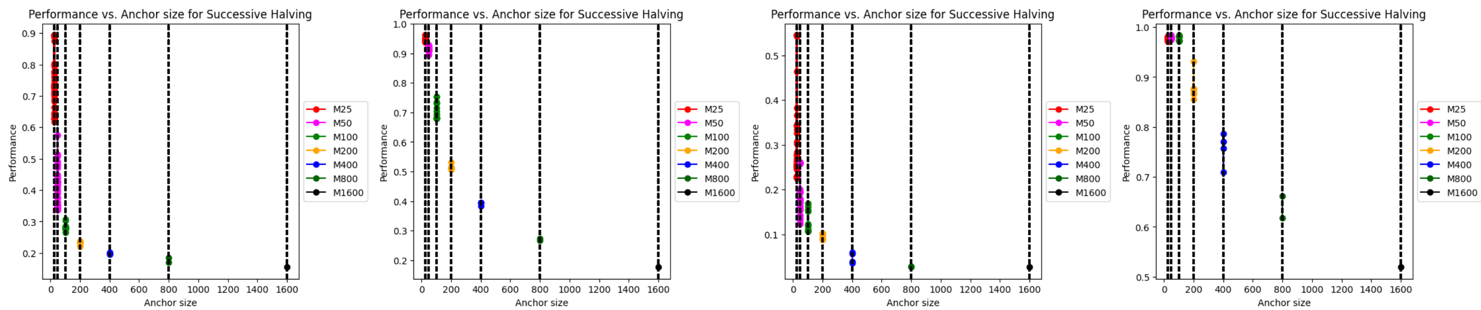
(a) Random Search (b) SMBO (c) Successive Halving

Figure 1: Working results of three optimization algorithms over 50 iterations on four different datasets. Only performances(error rate) with improvements are recorded. The three plots have approximately same results regardless of the minor differences.

2. The working of the Random Search, SMBO, and Successive Halving

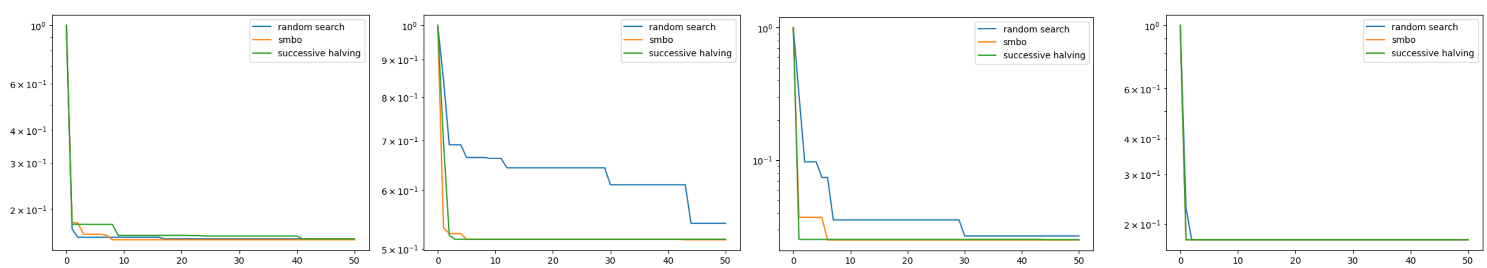
Figure 1 demonstrates the running results on different datasets using three algorithms. In this experiment, we use the full anchor size for Random Search and SMBO algorithms and choose the number of iteration to be 50 to speed up the process. With Random Search, the external surrogate model predicts performances of configurations by randomly selecting from the configuration space. SMBO with an internal surrogate model selects the next configuration to predict based on the value of an acquisition function. In Successive Halving, in each iteration, only the performance of the last remaining best configuration is recorded.

The last column in Table 1 shows the best performance results. On the lcd_configs dataset, SMBO achieves the best performance value. On the config-performances_dataset-6 dataset, the three algorithms approximately obtained the same results. While on the other two datasets, Successive Halving and SMBO find slightly better configurations than Random Search. Figure 2 demonstrates how Successive Halving performs in one iteration with different anchor sizes across four datasets.



lcdb_configs dataset-1457 dataset-11 dataset-6

Figure 2: The performance of Successive Halving across four datasets. Each point marked with different colors represents the performance of a configuration with different anchor size . Only the top-performing half of the configurations are allocated with more anchor size, finally the algorithm finds the best-performing configuration.



lcdb_configs dataset-1457 dataset-11 dataset-6

Figure 3: The comparison of the three methods on four different datasets. Successive Halving and SMBO converge fast across all datasets, while Random Search tends to perform less efficiently on dataset-1457 and dataset-11, particularly in earlier stages. On the first and last datasets, the gap between the three algorithms is very small.

3. Comparison of the three optimization algorithms on different datasets

   Figure 3 demonstrates the comparison of three optimization algorithms. SMBO, Successive Halving and Random Search reached almost the same results with iteration increases. While the primary difference lies in the first and the last datasets. We analyze potential reasons underlying this outcome. When dataset is small, the information gained by each algorithm varies minimally, which makes it not enough to reflect the impact of hyperparameters on performance and bring uncertainty to surrogate model, therefore resulting in similar performance of three algorithms[4]. The performance of surrogate model is generally more accurate when it is fitted on a large dataset and can better predict the performance of new hyperparameter configurations, weakening the relative advantage of optimizing algorithms[4].

References

[1] Jakob Bossek, Carola Doerr, and Pascal Kerschke. Initial design strategies and their effects on sequential model-based optimization: an exploratory case study based on bbob. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, GECCO '20. ACM, June 2020.

[2] Minsu Cho and Chinmay Hegde. Reducing the search space for hyperparameter optimization using group sparsity. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3627–3631, 2019.

[3] John Joshua Miller, Simon Mak, Benny Sun, Sai Ranjeet Narayanan, Suo Yang, Zongxuan Sun, Kenneth S Kim, and Chol-Bum Mike Kweon. Diverse expected improvement (dei): Diverse bayesian optimization of expensive computer simulators. *arXiv preprint arXiv:2410.01196*, 2024.

[4] Jeroen van Hoof and Joaquin Vanschoren. Hyperboost: Hyperparameter optimization by gradient boosting surrogate models, 2021.