Deep Learning Lab

Automated Algorithm
Design

Exercise 4

Author: Leslie Lydia Kurumundayil (4452075)

Date: 08.01.2019

Objective

The objective of this exercise to learn how Bayesian optimization and Hyperband work and how we combine them to optimize the hyperparameters of a convolutional neural network.

Architecture and Dataset

The model consists of three convolutional layers with RELU activations and batch norm and one fully connected layer. The dataset used in this exercise is CIFAR-10. The weights are optimized by using Adam optimizer. In this exercise, we do not train the original benchmark (CNN model), but we use the prediction of a surrogate benchmark, which is a regression model. This regression model is trained on a large set of randomly sampled hyperparameter configurations of the original benchmark.

Code Implementation

The code consists of three main parts. The first part includes the implementation of Bayesian optimization. Hyperband is implemented in the second part. In the last part of the code, Bayesian optimization is combined with Hyperband to optimize the hyperparameters of the convolutional neural network.

Bayesian Optimization

Initially, a configuration space (as shown in Table 4.1) is created.

Hyperparameter	**Range**
learning_rate	[-6,-1]
batch_size	[32,512]
n_filters_1	[4,10]
n_filters_2	[4,10]
n_filters_3	[4,10]

Table 4.1. Configuration Space for Hyperparameters

Before the BO loop is implemented, the BO components (a Gaussian process, expected improvement and an optimizer) are defined. In the main loop, the acquisition function is optimized, the objective function is evaluated, the dataset is augmented and the model is updated. We also keep track of the incumbent and runtime after each iteration and plot them against each other.

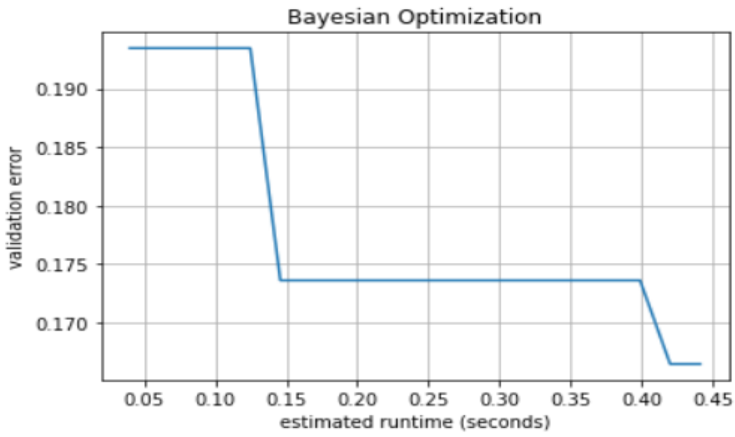


Figure 4.1. Bayesian Optimization

Hyperband

Hyperband uses random search with successive halving. Successive halving is a bandit-based technique for allocating more resources to promising configurations in a principled manner. The validation error is plotted against number of epochs, which is shown below.

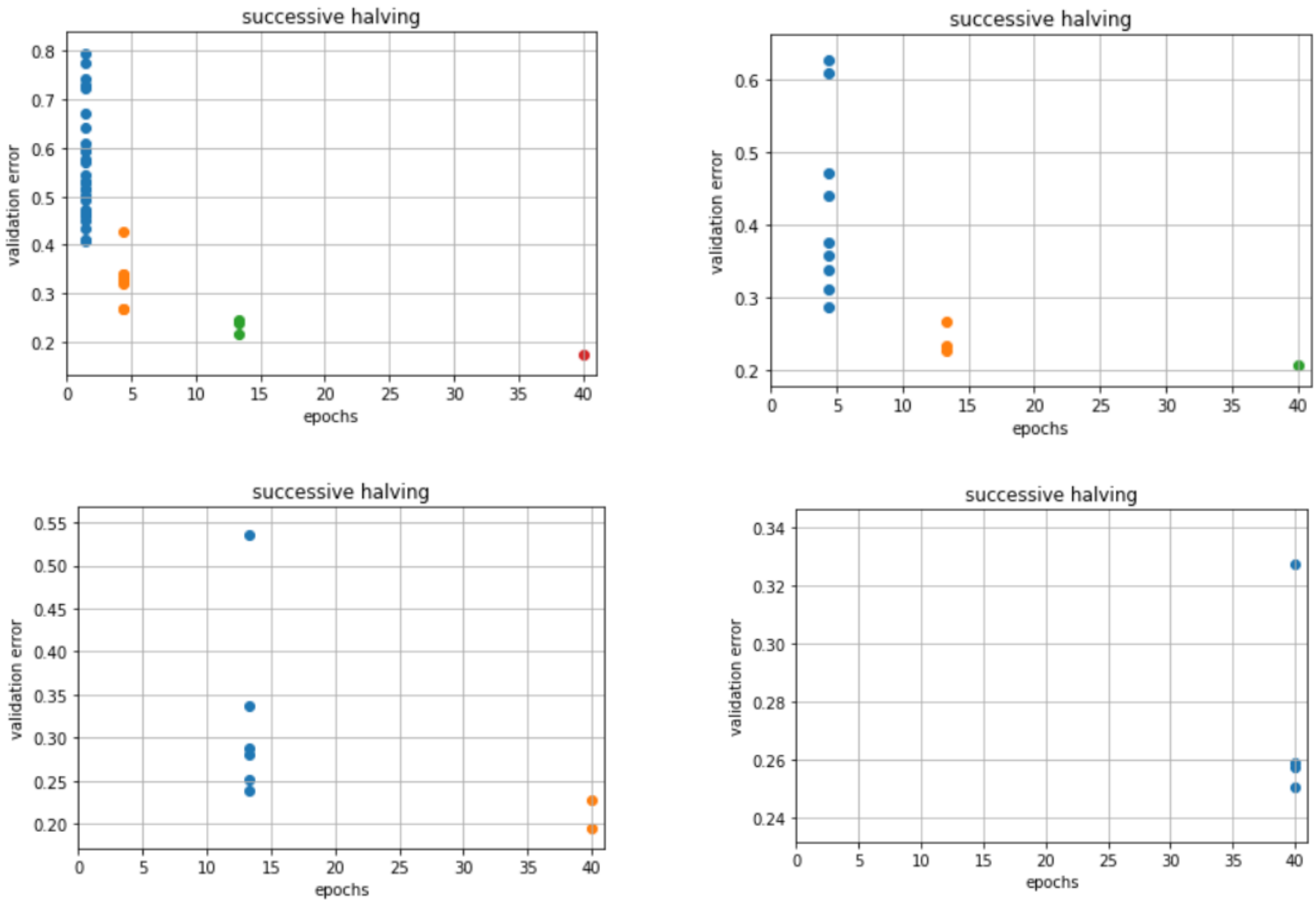
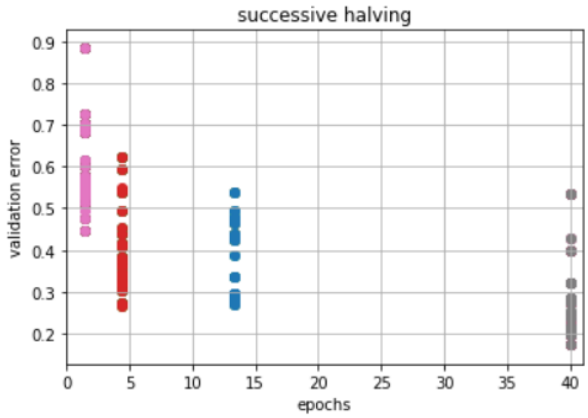
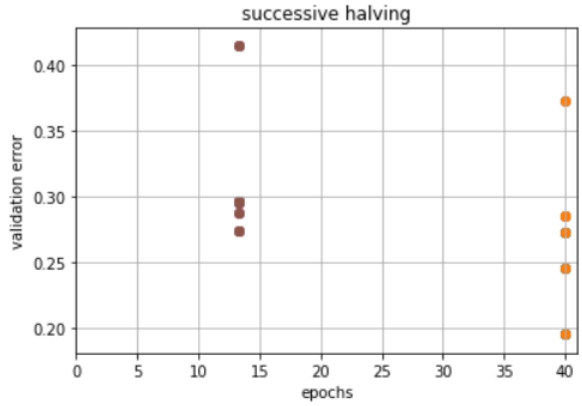


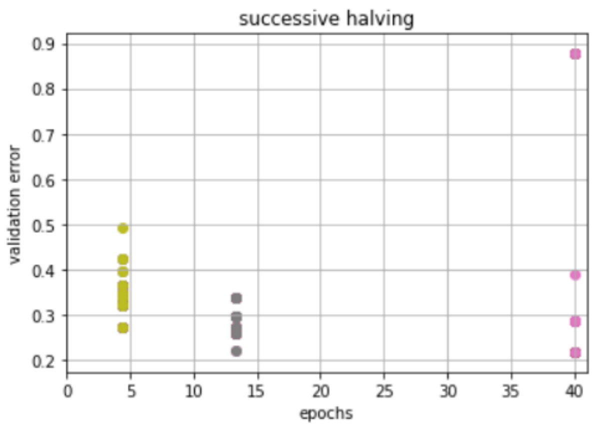
Figure 4.2. Hyperband (Validation Error vs Epochs)

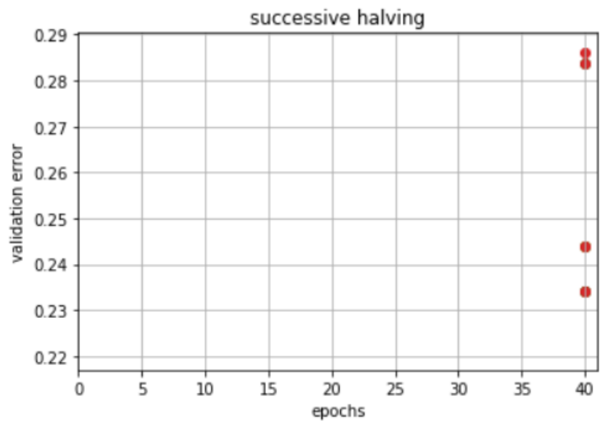
Combining Bayesian Optimization with Hyperband

The main disadvantage of Hyperband is that sampling of configurations is done at random which leads to an exponentially increased time to approach the global minimum. Thus, we combine Hyperband with a kernel density estimator that models the distribution of good and bad configurations. Instead of sampling from a uniform distribution, we sample from this model to find good configurations faster.









Results and Observations

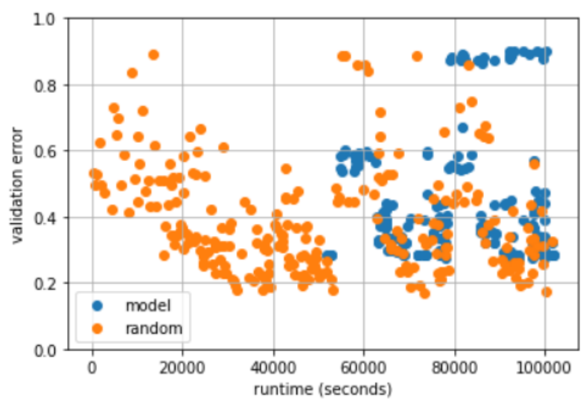


Figure 4.4. BOHB vs Random

Bayesian optimization with successive halving converges faster than the random.