Deep Learning Lab Report - Exercise 4

Sara Al-Rawi

Due: 08. January 2019

1 Introduction

In this lab exercise different hyperparameters optimization approaches have been investigated, in order to realize how a pure Bayesian Optimization have been improved by combining different approaches. To avoid long optimization time a so-called surrogate benchmark has been optimized instead of the original benchmark.

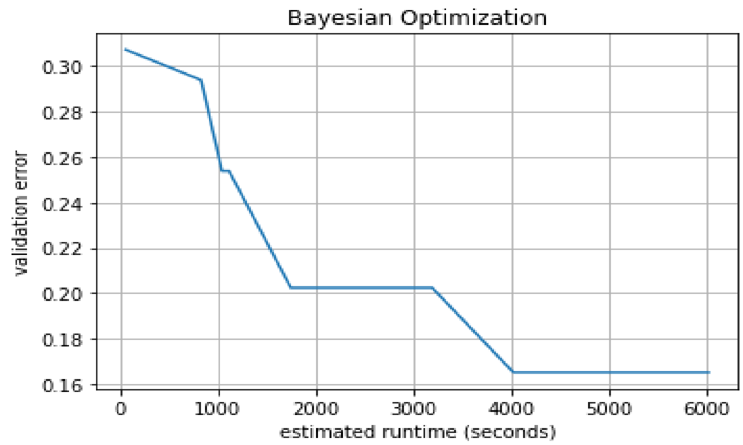
The lab work consists of three parts: **Bayesian optimization**, **Hyperband** and the last part which is **BOHB**, this approach combines the Bayesian optimization with Hyperband and yields the best results. The BO[1] is implemented using emukit.

2 The Implementation and Results

2.1 Bayesian Optimization

Since evaluating the real objective function is expensive, instead BO builds a surrogate (approximated objective function) for the objective and quantifies the uncertainty in that surrogate using a Bayesian machine learning technique, Gaussian process regression, and then uses an acquisition function defined from this surrogate to decide where to sample. In order to create the components of BO emukit has been used.
The figure below illustrates the validation error over the number of iterations.

[1]Bayesian optimization



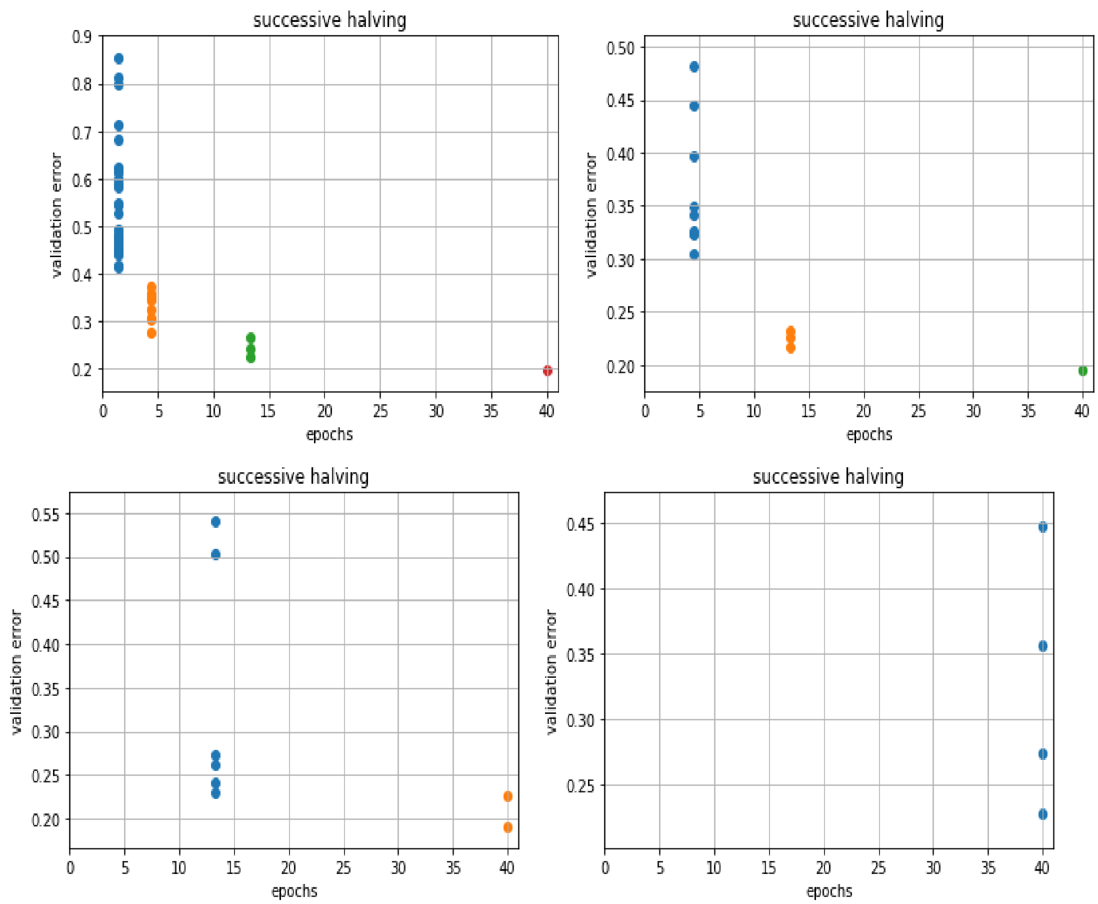
2.2 Hyperband

Hyperband is parameter-free and has strong theoretical guarantees for correctness and sample complexity. The approach relies on an early-stopping strategy for iterative algorithms of machine learning algorithms. The rate of convergence does not need to be known in advance and, in fact, the algorithm adapts to it so that if you replace your iterative algorithm with one that converges faster, the overall hyperparameter selection process is that much faster.

The underlying principle of the procedure exploits the intuition that if a hyperparameter configuration is destined to be the best after a large number of iterations, it is more likely than not to perform in the top half of configurations after a small number of iterations. That is, even if performance after a small number of iterations is very unrepresentative of the configurations absolute performance, its relative performance compared with many alternatives trained with the same number of iterations is roughly maintained. [2]
According to the above mentioned assumption, this part of the exercise has been implemented.

[2]reference: http://people.eecs.berkeley.edu/ kjamieson/hyperband.html

The graphs below have been obtained after running successive having for four unique executions and maximum number of epochs 40.



2.3 BOHB

One of the weaknesses of Hyperband is that it draws configurations randomly and hence might take exponentially long to approach the global optimum. BoHB combines Hyperband with a kernel density estimator that models the distribution of the good and the bad configurations in the input space. By sampling from this model instead of a uniform distribution the good can be found configurations much faster.

BOHB uses multivariate kernel density estimator to estimate both the good and bad distributions in order to fit all the dimensions and as a result capture interactions between hyperparameters. The implementation consists of sampling and updating the model.

The figure below illustrates that the samples that were drawn from our model performs better than the ones drawn randomly.