Deep Learning Lab

Autonomous Intelligent Systems

Exercise 4

Syed Sagheer Hussain (4353497)

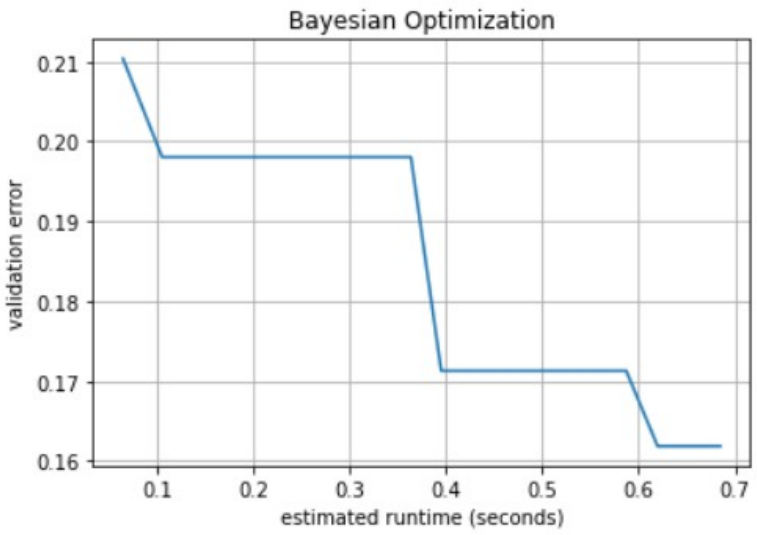
5.01.2019

Introduction

The focus of the exercise is Bayesian optimization and understanding of Hyperbdand. Eventually the combination of both for optimizing hyperparameters of a convolutional neural network (CNN). Instead of deploying the whole CNN directly, we have used given surrogate benchmark (regression-model) as random forest to evaluate the next hyperparameter configuration for evaluating the loss.

Bayesian Optimization

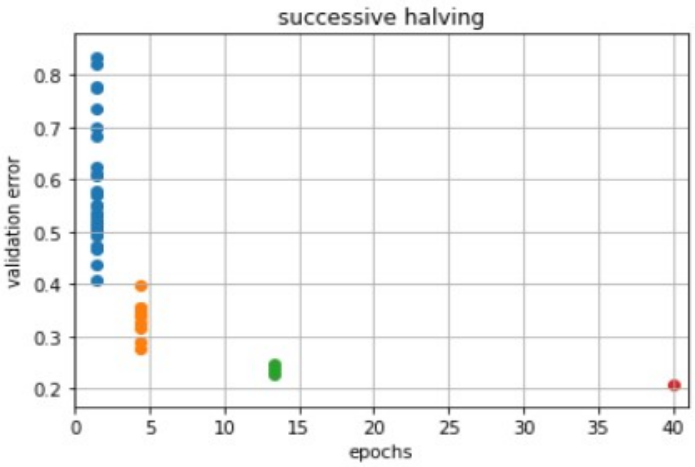
Hyperparameters required to optimize were the learning rate (on logarithmic scale), number of filters in each of the 3 convolutional layers, the batch size. All hyperparameters were considered as continuous variables and not discrete. Actually, the Bayesian optimization internally uses a model to guide the search. For train a model we have used initial design to collect data points (by N random configurations) before the Bayesian optimization loop. For estimation of real, we predict the training time of each hyperparameter configuration and plot the incumbent performance over the cumulative runtime.

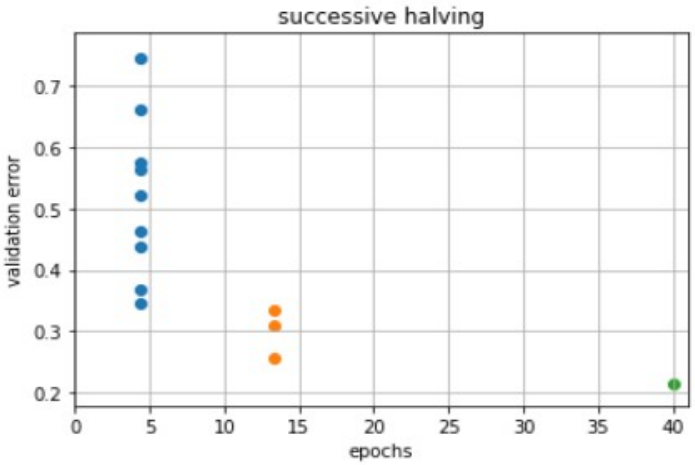


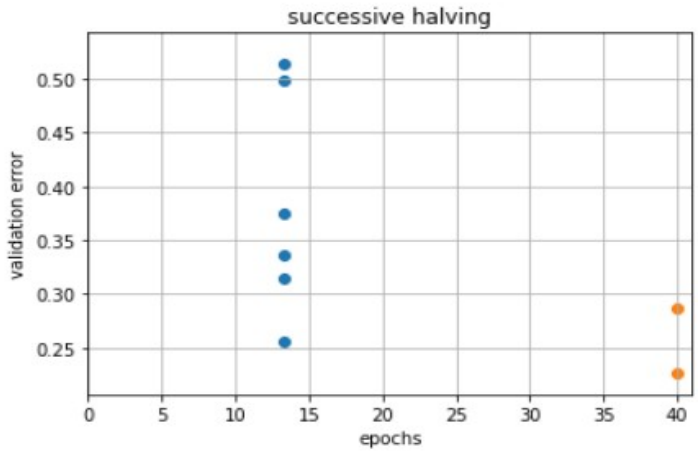
Hyperband

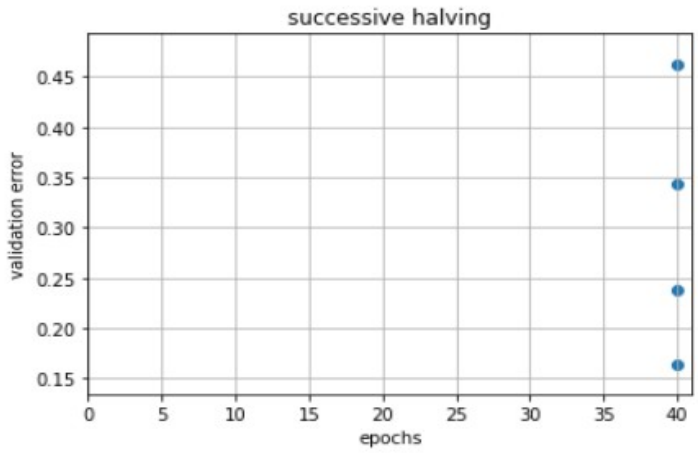
In approximation of the objective function, we have used the learning curve, i e. the validation error after each epoch. Hyperband combines random search with successive halving to balance very aggressive evaluation with many configurations on the smallest budget and very conservative runs on the maximum budget.

In hyperband we have sampled randomly some set of configurations of eta and B iterations. Inner loop describes the early-stopping while the outer one describes the hedge strategy. It proceeds considering multiple configurations in parallel and terminates poor performing configurations which eventually save more resources.


successive halving
validation error
epochs


successive halving
validation error
epochs


successive halving
validation error
epochs


successive halving
validation error
epochs

The last part was rather too complex for me. Understanding the given code as well as trying my own code result only in errors. I did not find the satisfying material on the internet either.