Optimising Bayesian optimisation for hyperparameter tuning

BoTorch

Charlie Hennys

Hyper-parameter optimisation

Neural networks have many hyperparameters that can affect performance (eg. learning rate, batch size, choice of optimiser).

Selecting these by hand is hard.

Basic methods such as grid search or random search don't work well when the cost of training the network is high.

Bayesian optimisation offers a smarter way of determining optimal hyperparameter configurations.

Bayesian Optimisation

We want to model the performance of a neural network as a function of its hyperparameters and find the maximum.

Bayesian optimisation has its own hyper parameters.

Choice of surrogate function

Choice of acquisition function

My project

Are default values provided by BoTorch (or other BO libraries) sufficient?

If not, is choosing the hyperparameters for BO by hand sufficient?

If not, should we perform hyperparameter optimisation on the BO hyperparameters?

One of the main challenges is that BO is already expensive so we don't want to run it multiple times. Can we use a coarser BO to reduce this cost? Can we transfer knowledge about BO parameters between models.

Plan

- No tuning
- Hand tuning
- Random search
- Grid search
- BO (default parameters)
- BO (hand selected parameters)
- BO (random/grid searched parameters)
- BO (optimised parameters)