Hi Nick,

I want to provide an intro on hyper-parameters training using Bayesian Optimization.

Step 1) Initial sampling set

Step 2) evaluations

Step 3) Training of gaussian process regressor

Step 4) Calculation of acquisition function

Step 5) Identify Input vector minimizing acquisition function value

It is an iterative process from step 5 to step 2 until optimal parameters are achieved.

Be aware of choosing locally best point.

grid search: Curate a list of all possible values for hyper-parameters and then try brute force method of trying all combinations.

Random search is a better option compared to grid search.

random search: pick one randomly from each hyper parameter list. Some parameters will have a lot of effect and others won't effect. Repeat those hyper parameters that have a lot of effect.

In a grid search, we have several points along each axis that are identical for all the other parameters, so if moving along that axis of the grid search makes no difference, we replicated same experiment many times and haven't learn anything.

Bayesian optimization is a machine learning approach to hyper parameter tuning.

It predicts regions of new hype parameters that will do better along with uncertainty.

Please reach out to me if you need any help

Thanks,

Rohit