

Hyperparameter optimization is also known as tuning; and is the problem of finding optimal hyperparameters for learning.

Machine Learning involves two types of params.

- hyperparams
- "normal" params.

The difference between the two is that hyperparams are set as a means of controlling the learning proc. whereas the others are "learned".

The same model's config - i.e constraints, weights, learning rates may change depending on the pattern of the data.

That config is the hyper-params. The params which most optimally "learn"/solve ML are what minimizes the loss function.

- Use cross-validation to generalize performance.

## Approaches

reason about quality  
of experiment

### Bayesian Optimization

#### GRID SEARCH

"an exhaustive search; sometimes constrained"

#### RANDOM SEARCH

"good for discrete and continuous spaces. good w- low dimension. can use prior knowledge of data distrib."

Probabilistic model of hyperparams and loss func.

Balance exploration of most uncertain params & exploitation of hyperparams close to the optimum.

## GRADIENT BASED OPTIMIZATION

When possible can compute gradient of the derivative.

Optimize hyp-params with gradient descent.

used NN, logistic reg. & SVMs.

Can also work analytically using automatic differentiation (AD)

↳ just do math & take derivative.

## OTHERS

- spectral and RBF

## EVOLUTIONARY OPTIMIZATION

use evolutionary algo to search for hyper params.

use fitness function -

like 10 fold cross-validation.

and rank/evolutionary Selection by fitness.

## POPULATION BASED?

Multiple learning instances that work collectively by evolution to optimize the loss function.