AML, 1 Oct 2019

Bayesian Optimization (Frazier 2018)

To find $\max_{x \in A} f(x)$

Evaluation of $f(x)$ extremely expensive

Neural network

Given a set of hyperparameters (# nodes, # layers etc) – fix an architecture

$f(x)$ = accuracy from model

How to determine optimal values for hyperparameters?

Use Bayesian optimization

Assume inputs $\in \mathbb{R}^d$ Works well for $d \leq 20$

Hyperparameters typically few – 5 or 6

$n = n_0$ – possible x (hyperparameters)

n_0 user defined, say 20

$x_1 \quad x_2 \quad \dots \quad x_{n_0}$

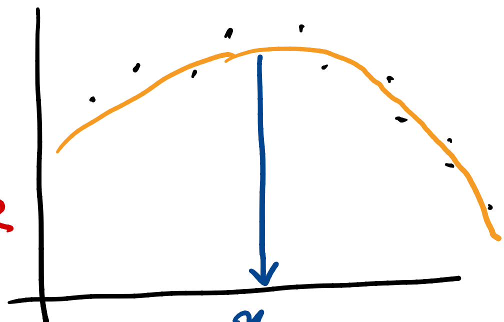
$NN(x_1) \quad \dots \quad NN(x_{n_0})$

$f(x_1) \quad \dots \quad f(x_{n_0}) \leftarrow$ Accuracy

$\quad \| \qquad\qquad\qquad \|$

$y_1 \quad \dots \quad y_{n_0}$

- Treat accuracy as target value
- Plot points and perform GPR
- Choose max as x_{n+1}



Extend sequence $NN(x_{n+1})$

$f(x_{n+1}) = y_{n+1}$

Recalculate surface, choose global optimum as x_{n+2}

Convergence ?

Acquistion function

$$f_n^* = \max_{m \leq n} f(x_m)$$

Expected improvement

$$EI_n(x) = E_n\left[\,|\,f(x) - f_n^*\,|\,\right]$$

Define $a^+ = \max(a, 0)$

$$EI_n(x) = [\Delta_n(x)]^+ + \sigma_n(x)\,\Phi\left(\frac{\Delta_n(x)}{\sigma_n(x)}\right) - |\Delta_n(x)|\,\Phi\left(\frac{\Delta_n(x)}{\sigma_n(x)}\right)$$

$\Delta_n(x) = \mu_n(x)\ f_n^*$

$\mu_n(x) =$ mean GP estimator $\{(x_i, y_i) \mid i \in 1..n\}$

$\sigma_n^2(x) =$ variance " " "

Other criteria can be used for convergence

scikit-learn has 3 options

Can apply Bayesian Optimization to other classification models, not only NN

[Snoek et al, 2012]