Bayesian Optimization

Hyperparameter optimization

$P_1, P_2, \ldots, P_k$ – each has a range of values

Grid Search

Enumerate combinations and choose the best

Computationally expensive – "brute force"

└ Can parallelize

Optimize the "performance"

- Costly to evaluate

$f(v_1, \ldots, v_n)$ v_i is setting for p_i

Cost function is a black box

- Kind of regression to "estimate" the shape of this function
- Normally, we make assumptions about type of function

Linear → Non linear?

Single variable case

$$x \longrightarrow \underbrace{\Big(\underset{a_0}{\overset{y_0}{1}},\ \overset{y_1}{x},\ \overset{y_2}{x^2},\ x^3,\ \ldots,\ \underset{a_{10}}{\overset{y_{10}}{x^{10}}}\Big)}_{\text{Linear regression here}}$$

polynomial function of x

More expansive version of this idea

Regression

Set of basis functions $\phi_1, \phi_2 \ldots, \phi_n$

Compute coefficients $\alpha_1, \alpha_2 \ldots \alpha_n$

$$f = \sum \alpha_i \phi_i$$

Fourier transform

Same idea

Infinite basis

Multivariate Gaussian

$f(x_1, \ldots, x_n)$

- Means
- Covariance matrix

Gaussian Process

Infinite dimensional version of a multivariate Gaussian

$(x_1, x_2 \ldots \ldots \quad)$

For any finite subset – multivariate Gaussian

The mean & covariance matrix is the <u>same</u> for every subset

A function of $(x_1, x_2, \ldots x_n)$

Gaussian Process Regression

Basis is a family of functions

Find the best fit

Bayesian Optimization at work

Start with $x_1, x_2, \ldots x_n$ (each x_i is a vector)

Evaluate my model at each x_i & compute $f(x_i)$

Fit a multivariate Gaussian to these observations

- Choose a "reasonable" format for mean & covariance

typically 0

↳ x_i & x_j are "close", covariance is higher

Given this $\hat{f}$ fitted to $(x_1, \ldots, x_n)$

└ Estimate $\hat{f}(x_{n+1})$ for any x_{n+1}

$\mu_{n+1}, \sigma^2_{n+1}$

Not re-running

ML model for x_{n+1}

Acquisition step

- Choose the "best" x_{n+1}

- Optimize over x_{n+1}'s based on regression prediction