Bayesian Optimization

# Bayesian Optimization

We have function f: with to minimize on some domain X . If **a functional form** for f is **not available,** we Bayesian Optimization proceeds by maintaining a probabilistic belief about f and designing an acquisition function to determine where to evaluate the function next.

Bayesian optimization almost always reason about f by choosing an appropriate **Gaussian Process prior**:

Given observation we can condition our distribution D to compute posterior expectation of the function f is look likes . How can select where to observe next? The acquisition function is inexpensive function that evaluated at a given point to measure how desirable evaluating is expected to be for minimization problem. We then can optimize the acquisition to select region of domain of f are optimal (location of next observation).

To compute posterior, we need a likelihood model for the samples from f and prior probability model on f. We can assume normal likelihood with noise

For the prior distribution, assume function f can be described by a Gaussian Process (GP). For data point we assume value of the function can be described by a multivariate Gaussian distribution

# Acquisition function

To find the best point to sample next, we need an objective function that is acquisition function. This is a

Probability of improvement

is the minimal value of f observed. PI evaluate f at the point most likely to improve on this value. Utility function associated with evaluating f at a given point x:

The probability of improvement acquisition function is expected utility as a function of x. The point with highest probability of improvement is selected

Expected improvement

It is similar with PI but it takes count the size of the improvement. EI evaluate f at the point in expectation most improvement. This corresponds to the following utility function

The expected improvement acquisition function then the expected utility as a function of x. The point with highest expected improvement is selected

EI has 2 components. The first can increase by reduce mean of function and the second can increase by increasing variance . These 2 terms can be interpreted as a tradeoff between **exploitation** (points with low means) and **exploration** (points with high uncertainty)

Entropy Search

We seek to **minimize the uncertainty** we have **in the location of the optimal value**. . ES seek to evaluate points so as to minimize the entropy of the induced distribution .

This is can be done by, first, computing current amount of information H about minimum. Second, approximate the expected information gain at certain location. Finally, suggesting next evaluation point where is maximize. Utility function at x

\*P/s: Amount of information about the location of minimum is computed

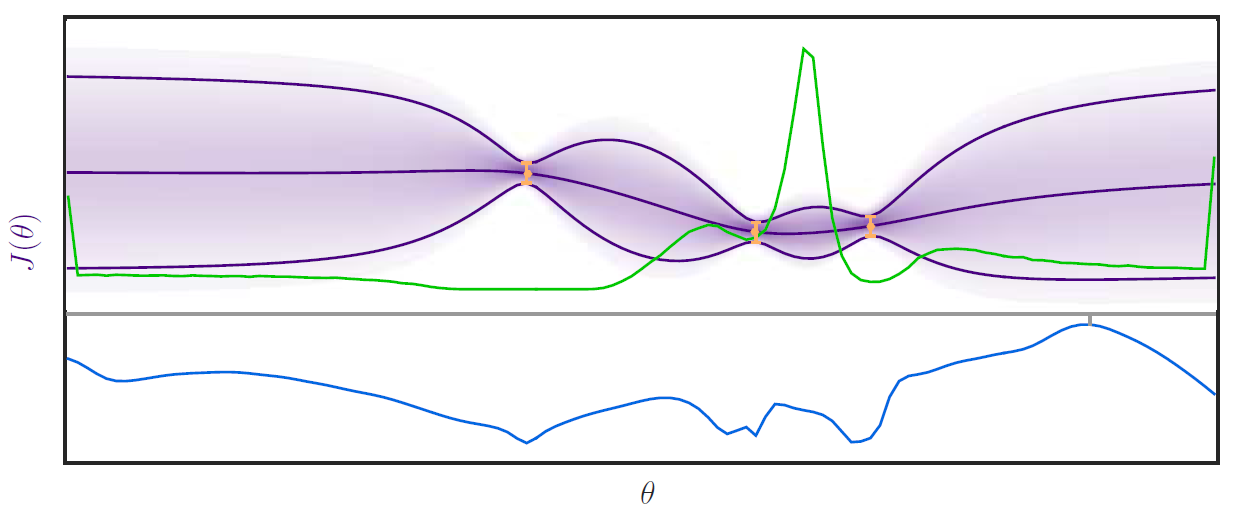


Figure Approximated probability distribution over the location of the minimum p\_min(Θ) in green and The blue line represents the expected gain in information E [ΔH] (Θ).

Our entropy search acquisition function then the expected utility as a function of x

Due to no closed-form expression for distribution of . A series of approximation must be made

Upper confidence bound

Acquisition function take the form