μ and σ are functions of x[1]

Last Meeting

We construct the mean vector by evaluating a *mean function* μ_0 at each x_n. We construct the covariance matrix by evaluating a *covariance function* or *kernel* Σ_0 at each pair of points x_i, x_j. The kernel is chosen so that points x_i, x_j that are closer in the input space have a large positive correlation, encoding the belief that they should have more similar function values than points that are far apart. The resulting prior distribution on $[f(x_1), ..., f(x_n)]$ is

$$\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_m) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_m) \end{bmatrix} \middle| \begin{bmatrix} \mu_0(x_1) \\ \mu_0(x_2) \\ \vdots \\ \mu_0(x_m) \end{bmatrix}, \begin{bmatrix} \Sigma_0(x_1,x_1)\ldots\Sigma_0(x_1,x_N) \\ \Sigma_0(x_2,x_1)\ldots\Sigma_0(x_2,x_N) \\ \vdots \\ \Sigma_0(x_N,x_1)\ldots\Sigma_0(x_N,x_N) \end{bmatrix}\right)$$

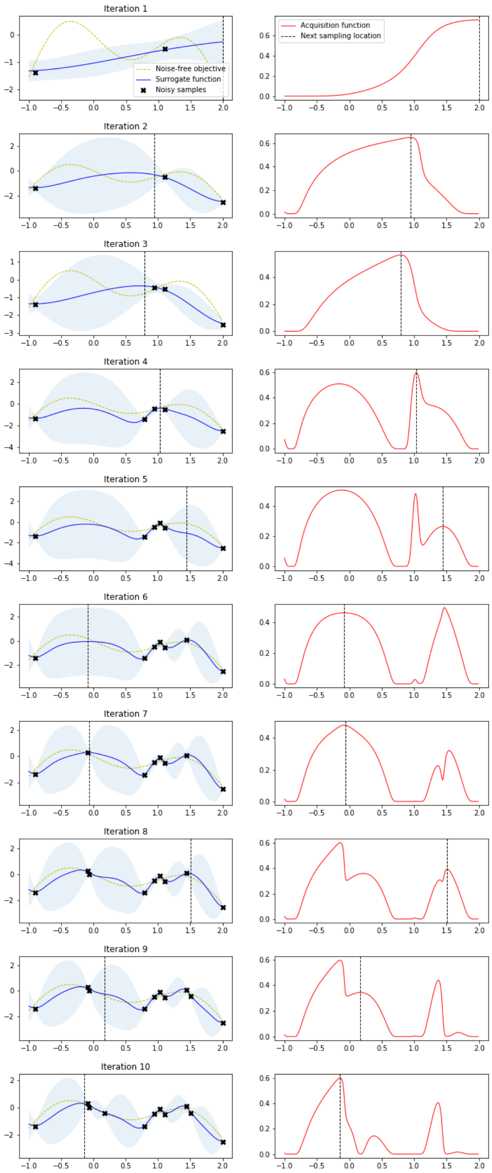
$$\boldsymbol{f}_N \sim \mathcal{N}(\boldsymbol{f}_N|\boldsymbol{\mu}_N, \boldsymbol{\Sigma}_N)$$

Suppose we observe $\boldsymbol{f}_N$ without noise for some N and we wish to infer the value of $f(x)$ at some new point x. To do so, we have the prior over $[\boldsymbol{f}_N, f(x)]$ is given by above equation. We can compute the conditional posterior of $f(x)$ using Bayes' rule and we get

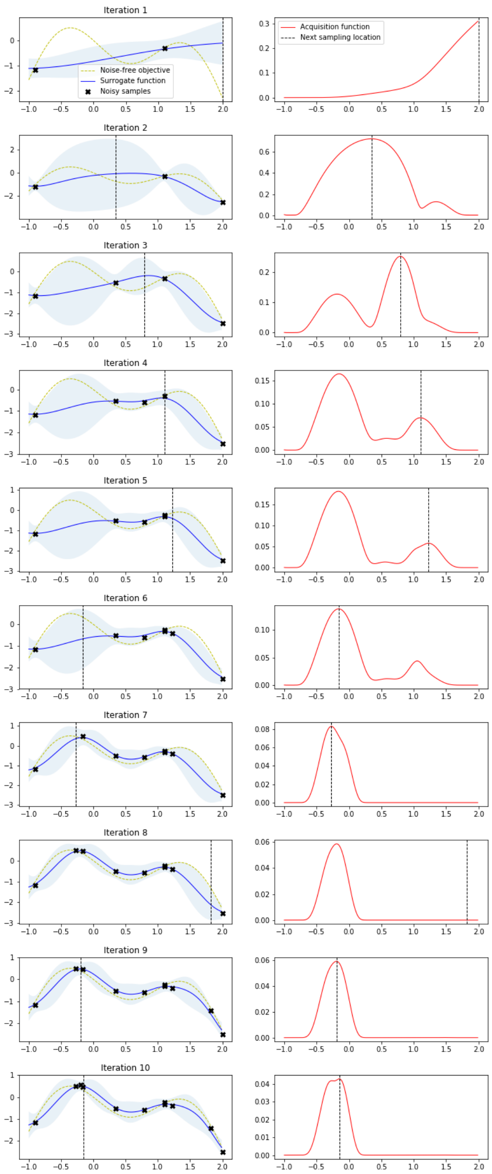
$$\begin{aligned} f(x)|\boldsymbol{f}_N &\sim \mathcal{N}\left(f(x)|\mu(x), \sigma^2(x)\right) \\ \mu(x) &= \boldsymbol{\Sigma}_*\boldsymbol{\Sigma}_N^{-1}(\boldsymbol{f}_N - \boldsymbol{\mu}_0) + \mu_0(x) \\ \sigma^2(x) &= \Sigma_0(x,x) - \boldsymbol{\Sigma}_*\boldsymbol{\Sigma}_N^{-1}\boldsymbol{\Sigma}_*^\top \\ \boldsymbol{\Sigma}_* &= \left[\Sigma_0(x,x_1)\ \Sigma_0(x,x_2)\ldots\Sigma_0(x,x_N)\right] \end{aligned}$$

Bayesian Optimization Plots

Probability of Improvement



Expected Improvement



References

[1] Eric Brochu, Vlad M Cora, and Nando De Freitas. “A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning”. In: *arXiv preprint arXiv:1012.2599* (2010).