Lecture 23: Bayesian global optimization

Professor Ilias Bilionis

Probability of improvement - No observation noise

Derivation of the probability of improvement

$$x_{1:n} = (x_1, \ldots, x_n); \quad y_{1:n} = (y_1, \ldots, y_n); \quad \sigma = 0$$

$$f(\cdot) \sim GP(m(\cdot), c(\cdot, \cdot)) \Rightarrow \text{posterior point predictive distribution}$$

$$p(y \mid x, x_{1:n}, y_{1:n}) = N\left(y \mid m_n^*(x), \sigma_n^{*2}(x)\right)$$

$$\text{Current best observed output:} \quad \bar{y}_n = \max_{1 \le i \le n} y_i$$

$$a(x) = P\left(y > \bar{y}_n + \psi \mid x, x_{1:n}, y_{1:n}\right) = \int^{+\infty}_{\bar{y}_n + \psi} \underline{p(y \mid x, x_{1:n}, y_{1:n})} \, dy$$

$$= \left[\begin{array}{c} y \mid x, x_{1:n}, y_{1:n} \sim N(\ldots) \\ z = \dfrac{y - m_n^*(x)}{\sigma_n^*(x)} \sim N(0, 1) \end{array} \right] = P\left(z > \dfrac{\bar{y}_n + \psi - m_n^*(x)}{\sigma_n^*(x)}\right)$$

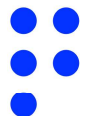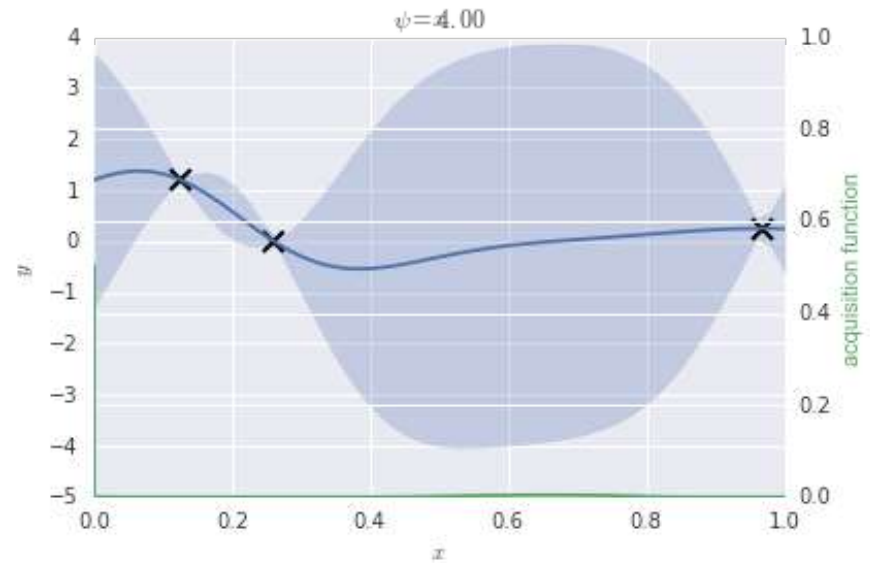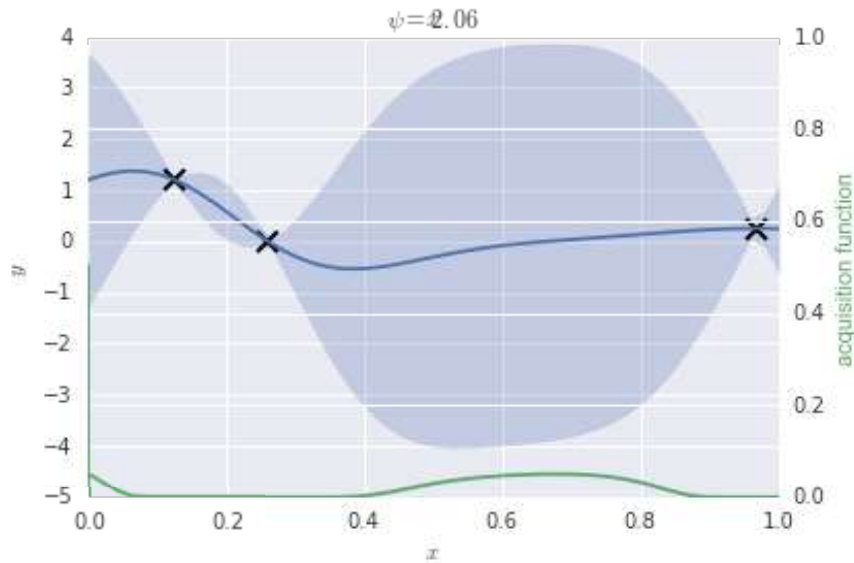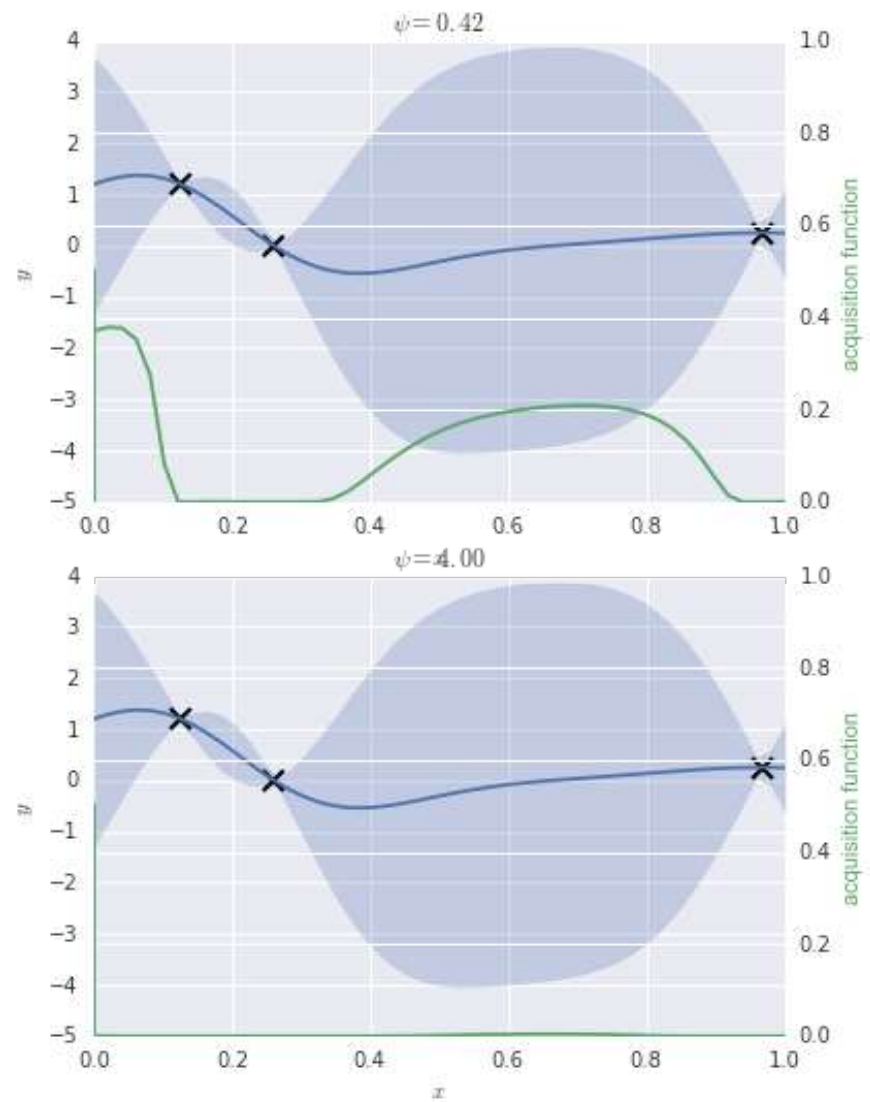$$= 1 - \underset{\text{CDF of } N(0,1)}{\underbrace{\Phi}}\left(\dfrac{\bar{y}_n + \psi - m_n^*(x)}{\sigma_n^*(x)}\right)$$

$$= \Phi\left(\dfrac{\bar{y}_n + \psi - m_n^*(x)}{\sigma_n^*(x)}\right)$$

30

Probability of Improvement
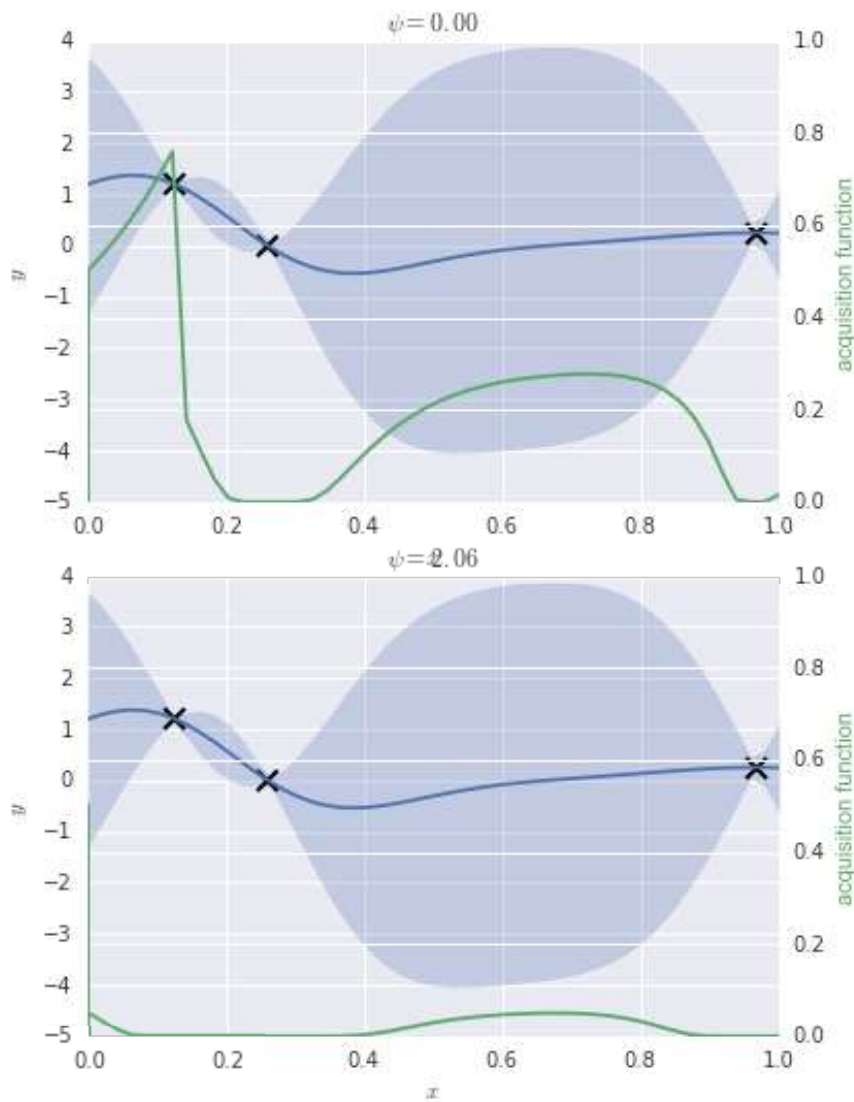
Probability of Improvement - Why not use it?

- Large value of ψ -> exploration.

- Small value of ψ -> exploitation.

- But how to you pick it?