贝叶斯优化 $\max f$

$f \mid X, y$

$$f(x) \sim N(\mu_n(x), \sigma_n(x)^2)$$

$$\mu_n(x) = [k(x, x_1) \cdots k(x, x_n)] K^{-1} \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{pmatrix}$$

$$\sigma_n(x)^2 = k(x, x) - [k(x, x_1) \cdots k(x, x_n)] K^{-1} \begin{pmatrix} k(x, x_1) \\ \vdots \\ k(x, x_n) \end{pmatrix}$$

采样 $\mu_n(x)$ 较大，$\sigma_n(x)$ 较大的点 不确定性大，需要探索

exploitation exploration

① 信息增益

$$y_i = f(x_i) + \varepsilon \quad i = 1, 2 \cdots n$$

$$\varepsilon \sim N(0, \sigma^2)$$

$$I(y; f) = \int p(y, f) \log \frac{p(y, f)}{p(y) p(f)} \, dy \, df$$

若独立，无信息增益

$$= \int p(y, f) \log \frac{1}{p(y)} + p(y, f) \log p(y \mid f) \, dy \, df$$

$$= \underbrace{\int p(y) \log \frac{1}{p(y)} \, dy}_{H(y)} - \underbrace{\int p(y, f) \log p(y \mid f) \, dy \, df}_{H(y \mid f)}$$

例子 f 今天云的情况，y 明天下雨的情况

$H(p)$：信息熵 binary $H(p) = -p \log p - (1-p) \log (1-p)$

$p(y) \sim N(0, K_{nn} + \sigma^2 I)$

$p(y|f) = N(0, \sigma^2 I)$ 与 f 无关

$$H(N(\mu, \Sigma)) = \int N(x;\mu,\Sigma)\left(\frac{1}{2}\log|2\pi\Sigma| + \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

$$= \frac{1}{2}\log|2\pi\Sigma| + \frac{1}{2}n$$

$$= \frac{1}{2}\log|2\pi e\Sigma|$$

选取 x_{n+1}，最大化

$$I(y;f|X) = \frac{1}{2}\log|\sigma^{-2}K_{n+1} + I|$$

$$K_{n+1} = \begin{bmatrix} K_n & k^T \\ k^T & d \end{bmatrix} \qquad K_{n+1} + \sigma^2 I = \tilde{K}_{n+1}$$

$$k^T = [k(x_{n+1}, x_1) \quad k(x_{n+1}, x_2) \cdots k(x_{n+1}, x_n)]$$

$$d = k(x_{n+1}, x_{n+1})$$

$$\tilde{K}_{n+1} = \underbrace{\begin{bmatrix} I & 0 \\ k^T\tilde{K}_n^{-1} & 1 \end{bmatrix}}_{L}\underbrace{\begin{bmatrix} \tilde{K}_n & 0 \\ 0 & d - k^T\tilde{K}_n k \end{bmatrix}}_{S}\underbrace{\begin{bmatrix} I & \tilde{K}_n^{-1}k \\ 0 & I \end{bmatrix}}_{L^T}$$

$$I(y;f|X) = \frac{1}{2}\log|\sigma^{-2}LSL^T|$$

$$= \frac{1}{2}\log|\sigma^{-2}S|$$

$\Leftrightarrow$ 最大化 $d - k^T\tilde{K}_n k$

$\Leftrightarrow$ 最大化 $\sigma_n(x)$

② 上置信界采样函数

收敛性证明

$$\gamma_n = \max_{X \subset D:\, |X| = n} I(y_n; f_n) \qquad \beta_i = 2\log\left(|D|\, i^2 \pi^2 / 6\delta\right)$$

参见 washington lecture 13

③ 概率改进采样函数

$$f(x) \sim N(\mu(x), \sigma(x))$$

$$P(f(x) > f_n^*) = \int_{f_n^*}^{+\infty} N(f; \mu(x), \sigma(x))\, df$$

选取 $f(x)$ 比 f_n^* 大，概率最大的点

④ 期望改进采样函数

$$[f(x) - f_n^*]^+ := \max\{0, f(x) - f_n^*\}$$

$$f(x) \sim N(\mu_n(x), \sigma_n(x))$$

$$\int (f - f_n^*)^+ N(f; \mu_n, \sigma_n^2)\, df \qquad \left(Z_n = \frac{f_n^* - \mu_n}{\sigma_n}\right)$$

$$= \sigma_n \int (Z - Z_n)^+ N(Z, 0, 1)\, dZ \qquad \left(Z = \frac{f - \mu_n}{\sigma_n}\right)$$

$$= \sigma_n \int_{Z_n}^{\infty} (Z - Z_n)\, N(Z, 0, 1)\, dZ$$

$$= -Z_n \sigma_n \int_{Z_n}^{\infty} N(Z; 0, 1)\, dZ + \sigma_n \int_{Z_n}^{\infty} \frac{Z}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}}\, dZ$$

$$= -Z_n \sigma_n \int_{Z_n}^{\infty} N(Z; 0, 1)\, dZ + \frac{\sigma_n}{\sqrt{2\pi}} e^{-\frac{Z_n^2}{2}}$$

$$= (\mu_n - f_n^*)\,[1 - \Phi(Z_n)] + \sigma_n \phi(Z_n)$$

$$= (\mu_n - f_n^*)\,\Phi(-Z_n) + \sigma_n \phi(-Z_n)$$

$$= \Delta_n \Phi\left(\frac{\Delta_n}{\sigma_n}\right) + \sigma_n \phi\left(\frac{\Delta_n}{\sigma_n}\right)$$

最大 $\Delta_n \uparrow$ $\sigma_n \uparrow$ 相互平衡

⑤ 知识梯度采样函数

$$a_{KG}(x) \approx \mu_{n+1}^{*}(x; y_{n+1}) - \mu_n^{*}$$

其中 $y_{n+1} \sim N(f; \mu_n(x), \sigma_n(x))$