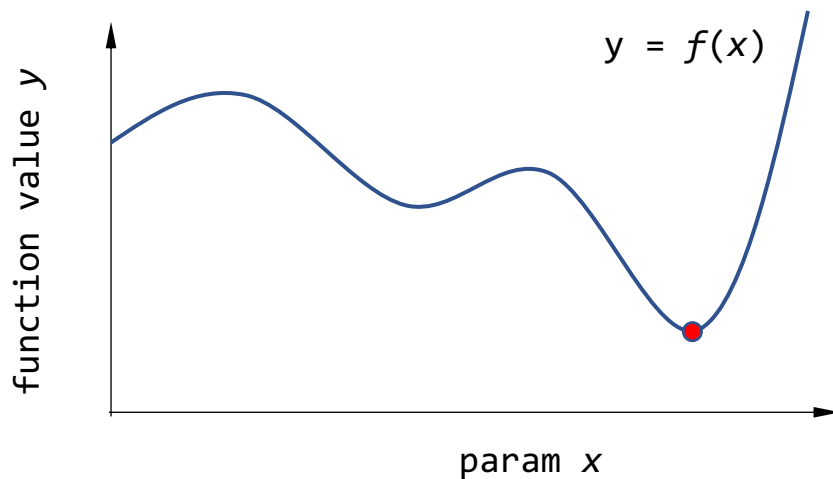


贝叶斯优化原理

罗磊 2019-04-08

一、目标问题



在封闭区间 $[x_{\min}, x_{\max}]$ 上求取以下最优问题：

$$x^* = \operatorname{argmin}_x f(x)$$

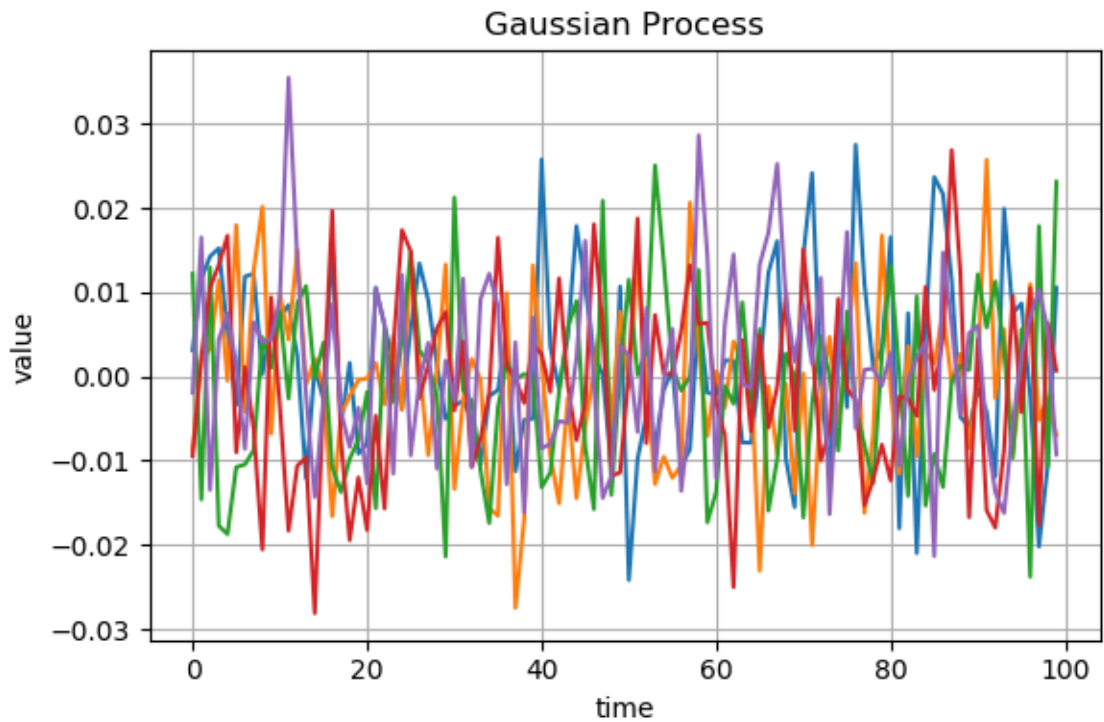
或

$$x^* = \operatorname{argmax}_x f(x)$$

其中自变量 x 可以是一维标量也可以是向量。

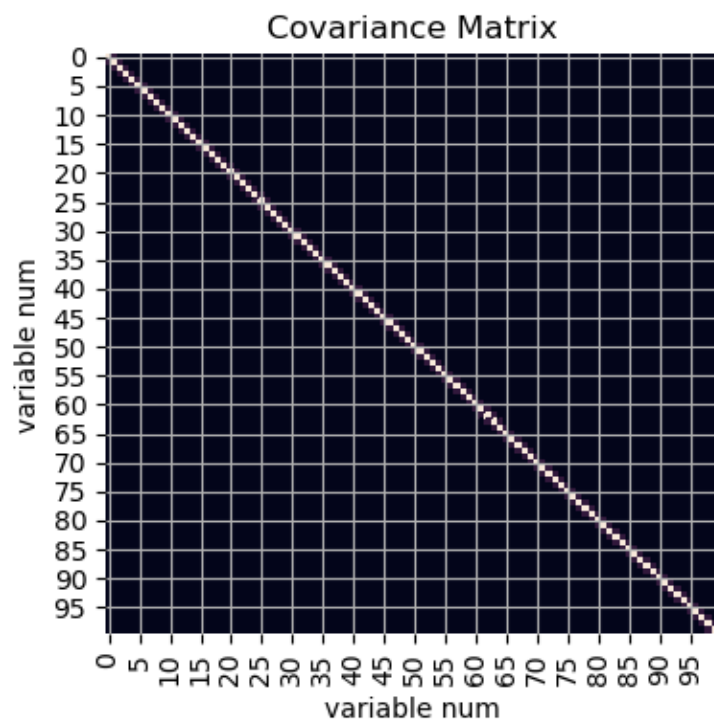
二、高斯过程与高斯核

高斯过程指的是一组随机变量的集合，这个集合里面的任意有限个随机变量都服从联合正态分布。

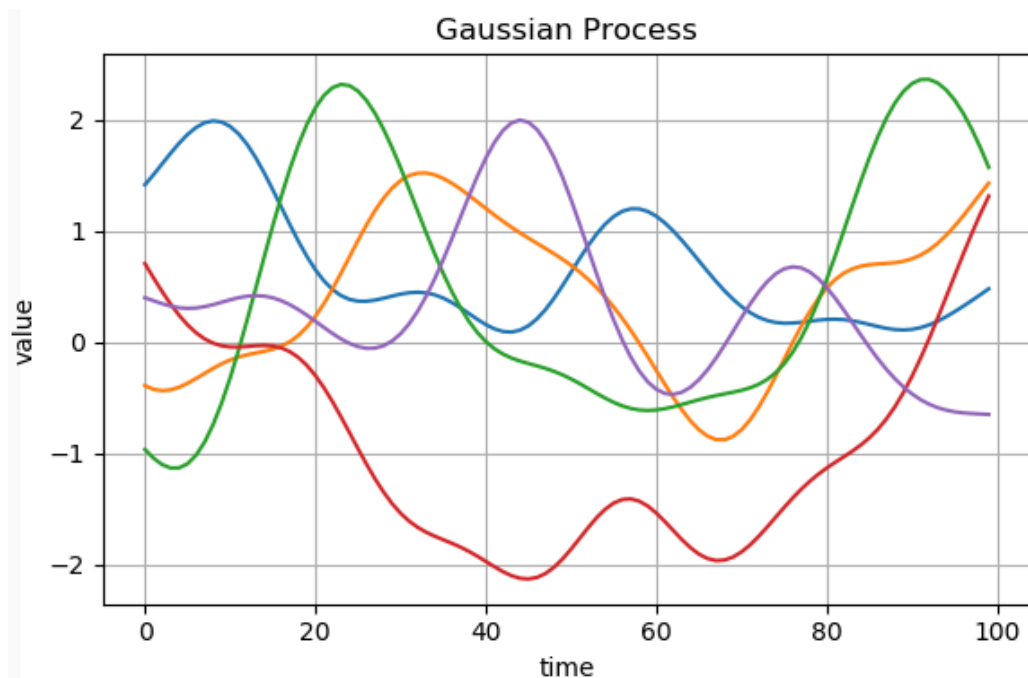
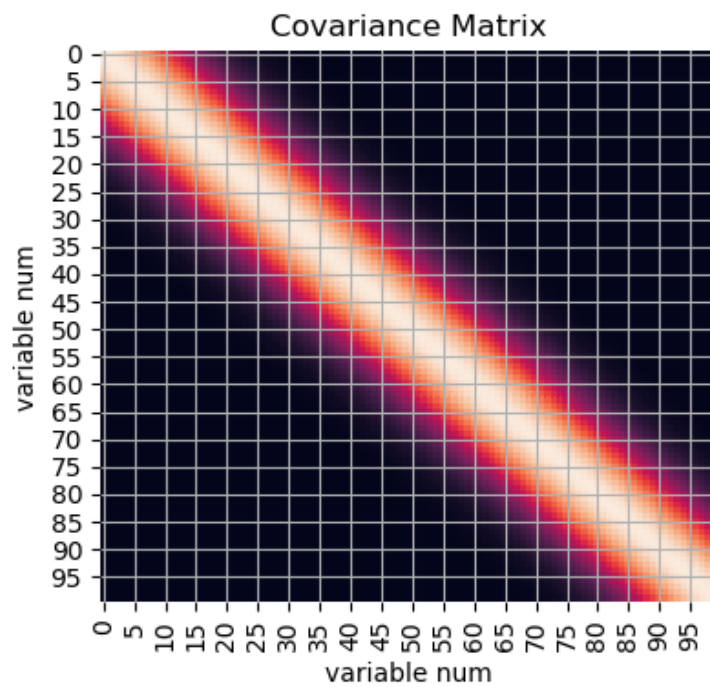


一个典型的高斯过程如上所示，不同颜色表示选取不同初始点获得的取样结果，高斯过程满足同一变量沿时间的取值符合高斯分布。

高斯过程中的变量在不同时刻的取值通过协方差矩阵进行关联，上例中各时刻取值对应的协方差矩阵如下所示：



可以看出，在上例中，相邻时刻间的协方差值较低，说明变量间关联性较弱，导致高斯过程曲线平滑性较差，若要提高曲线平滑性，可以考虑改变协方差矩阵值，提高相邻时刻间协方差关联度，如下所示：



理论上在该区间任意给出一连续函数均可以被高斯过程采样获得，

更准确地，至少能在有限时间内获得近似。下面我们就来讨论如何使用高斯过程结合贝叶斯原理对函数进行逼近和寻优。

三、函数估计

在区间 $[x_{min}, x_{max}]$ 上的任意连续函数可以视作上述高斯过程曲线中的某一支，此时横坐标由时间 **time** 变为 $[x_{min}, x_{max}]$ 上参数的所有可能取值 ($N + 1$ 维向量 x_s)。

$$x_s = [x_{min}, x_{min} + \Delta x, \dots, x_{max} - \Delta t, x_{max}]$$

对应的各维度均值为：

$$\begin{bmatrix} \mu_0 \\ \mu_1 \\ \vdots \\ \mu_N \end{bmatrix}$$

协方差矩阵可以通过高斯核函数计算，核函数可以如下形式：

$$kernal(x, x') = \exp\left(-\frac{(x - x')^2}{2}\right)$$

然后获得如下矩阵：

$$K = \begin{bmatrix} cov_0_0 & cov_0_1 & \dots & cov_0_N \\ cov_1_0 & cov_1_1 & \dots & cov_1_N \\ \vdots & \vdots & \ddots & \vdots \\ cov_N_0 & cov_N_1 & \dots & cov_N_N \end{bmatrix}$$

如果我们在 m 个参数位置上 x_{obs} 得到了 m 个计算结果 y_{obs} ，在不知道观测值得情况下，写出先验分布为：

$$\begin{pmatrix} y \\ y_{obs} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_0 \\ \vdots \\ \mu_N \\ \mu_{obs,0} \\ \vdots \\ \mu_{obs,m} \end{pmatrix}, \begin{pmatrix} K & K_* \\ K_*^T & K_{**} \end{pmatrix} \right)$$

其中：

$$\mathbf{K} = \text{kernal}(x_s, x_s)$$

$$\mathbf{K}_* = \text{kernal}(x_s, x_{obs})$$

$$\mathbf{K}_{**} = \text{kernal}(x_{obs}, x_{obs})$$

然后根据测量值，计算后验概率和最大似然估计，得到均值和方差的迭代更新公式：

$$\mu_s = \bar{x}_s + \mathbf{K}_*^T \mathbf{K}^{-1} (y_{obs} - \bar{x}_{obs})$$

$$\sigma_s = \mathbf{K}_{**} - \mathbf{K}_*^T \mathbf{K}^{-1} \mathbf{K}_s$$

如此不断进行迭代便可对未知函数进行估计。

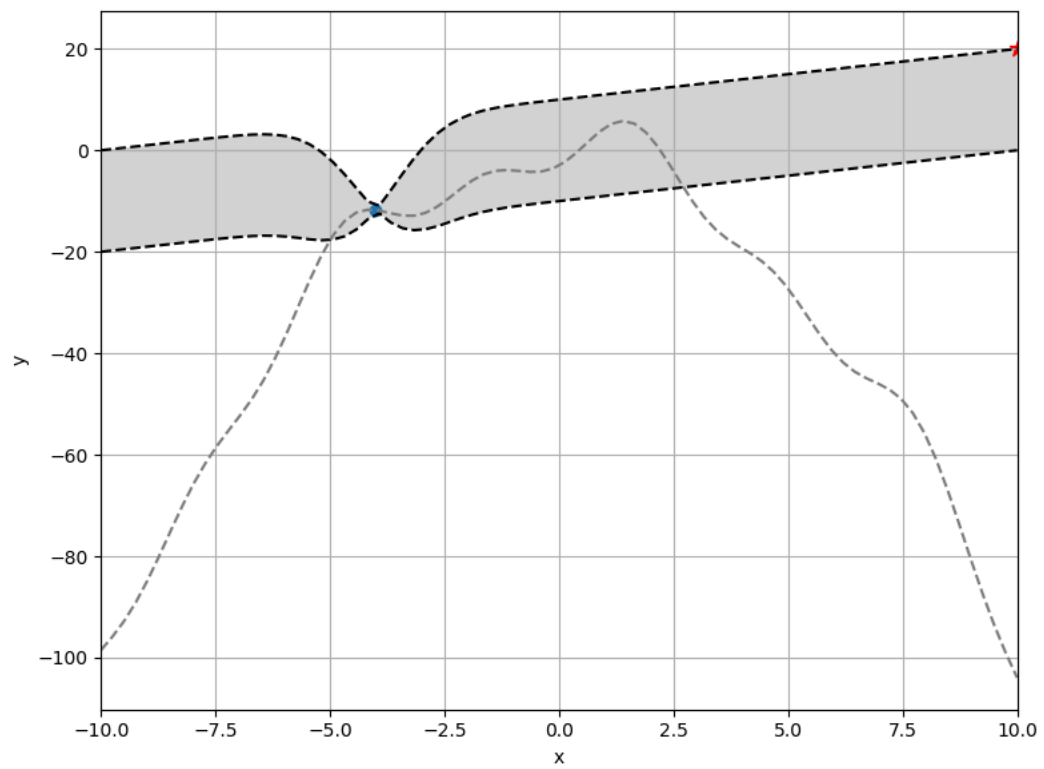
四、参数寻优

第三节中我们找到了一种通过高斯过程和贝叶斯估计对未知函数进行逼近的办法，实际应用中，我们并不需要对整段函数进行逼近，只需要获得函数最优值大小和对应的参数位置。此时我们可以考虑在上述高斯过程逼近的每一轮迭代中：

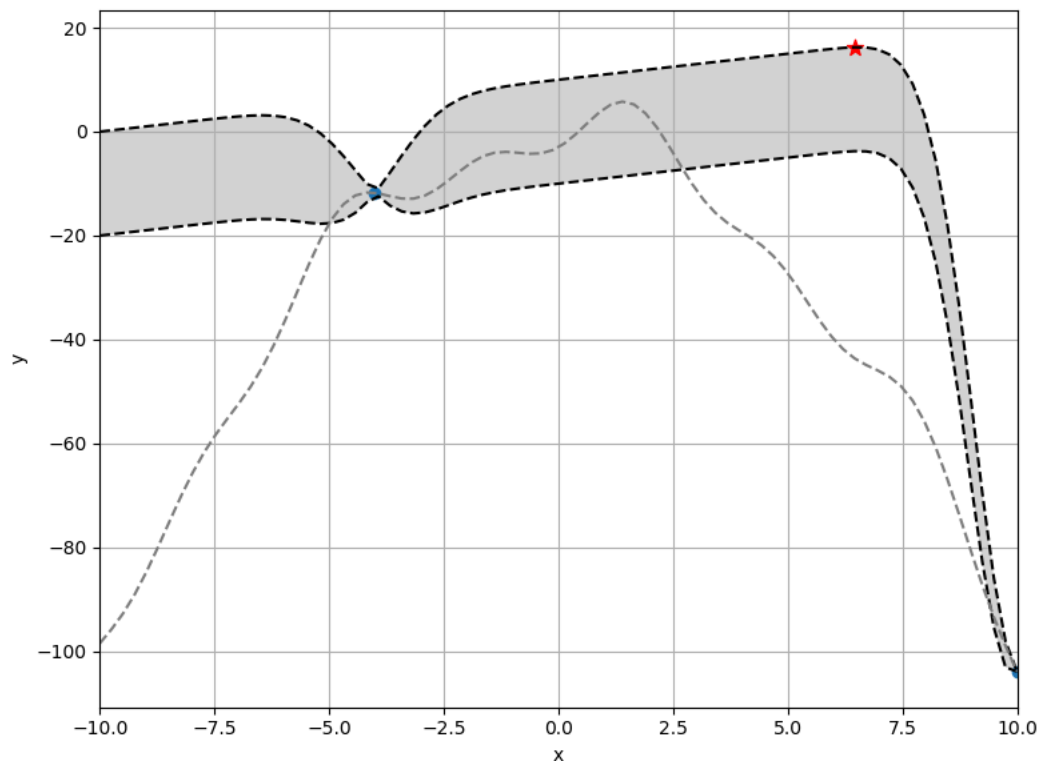
1. 记录更新后的均值 μ_s 和方差 σ_s ；
2. 根据这两个参数确定所有高斯曲线的大致分布，寻找出最高点位置；
3. 对寻找出的最高点位置进行采样，验证是否已经取得最大值；

寻优过程如下所示：

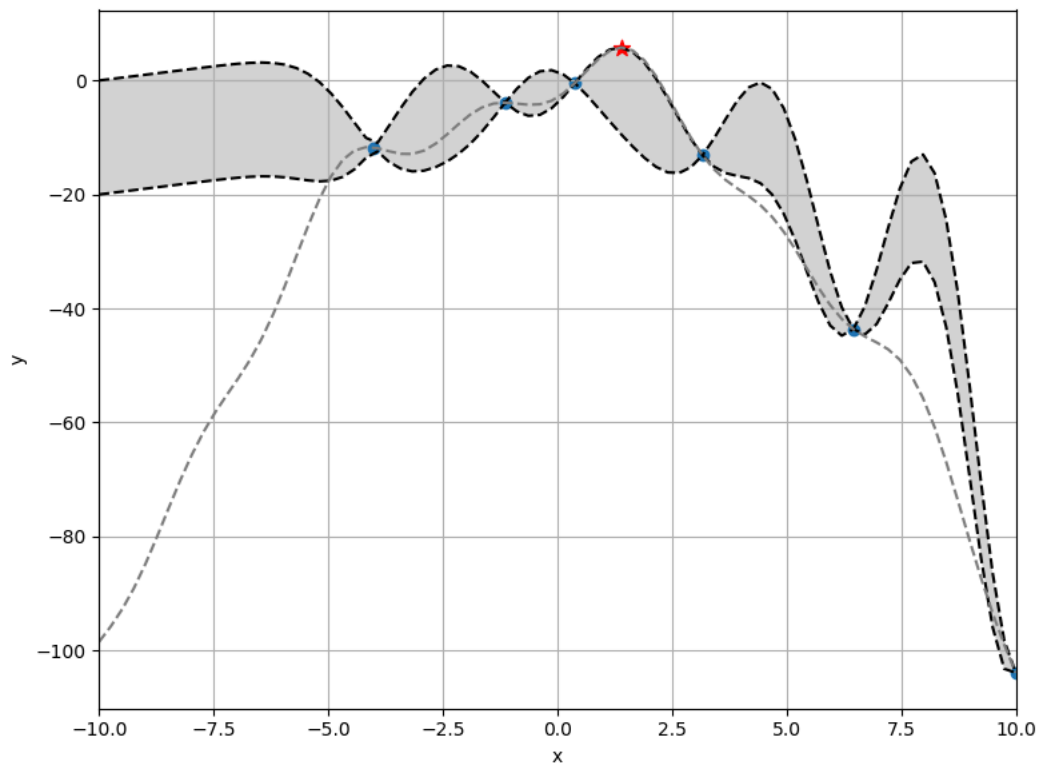
step = 0:



step = 1:



step = 5:



step = 10:

