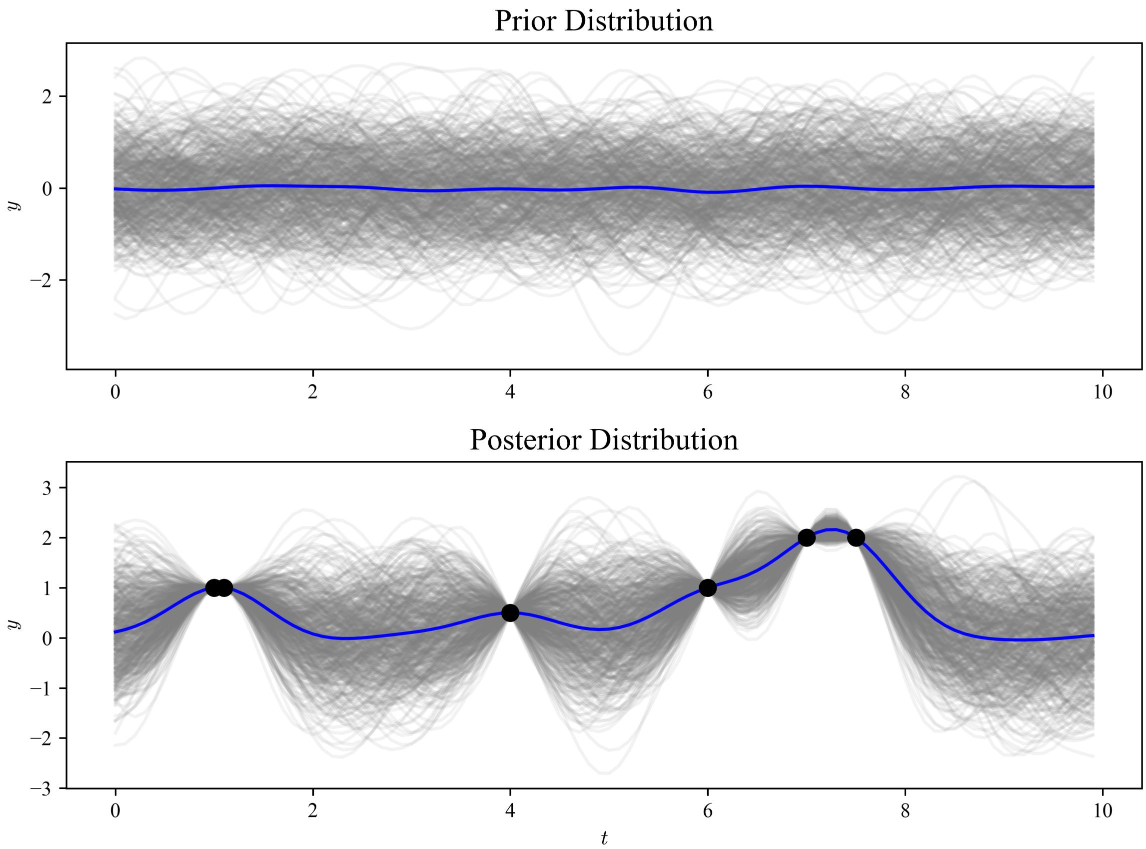


高斯过程回归

Gaussian Process Regression

一、高斯分布

高斯过程（Gaussian Process, GP）是随机过程之一，是一系列符合正态分布的随机变量在一指数集（index set）内的集合。该解释中的“指数”可以理解为“维度“，按照机器学习的角度，各个指数上的随机变量可以对应地理解为各个维度上的特征。

1.1 一元高斯分布

$$X \sim N(\mu, \sigma^2)$$

其概率密度函数为：

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-(x-\mu)^2/(2\sigma^2)\right)$$

标准正态分布：

$$\mu = 0, \sigma = 1$$

正态分布具有如下性质：

1. 如果$X \sim N(\mu, \sigma^2)$，且a、b均为实数，则$aX + b \sim N(a\mu + b, (a\sigma)^2)$；
2. 如果$X \sim N(\mu_x, \sigma_x^2)$与$Y \sim N(\mu_y, \sigma_y^2)$独立，则：
    1. $U = X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$；

   2. $V = X - Y \sim N(\mu_x - \mu_y, \sigma_x^2 + \sigma_y^2)$;
3. 若以上X与Y相互独立，则：
   1. XY符合以下概率密度分布：

$$p(z) = \frac{1}{\pi\sigma_x\sigma_y} K_0(\frac{|z|}{\sigma_x\sigma_y})$$

      其中K_0为修正贝塞尔函数；
   2. X/Y符合柯西分布：

$$X/Y \sim \text{Cauchy}(0, \sigma_x/\sigma_y)$$

4. 若$X_1, \ldots, X_n$各自独立，符合正态分布，则$X_1^2 + X_2^2 + \ldots + X_n^2$符合自由度为$n$的卡方分布；

1.2 二元高斯分布

$$f(x,y) = A\exp\left(-\frac{(x-x_0)^2}{2\sigma_x^2} - \frac{(y-y_0)^2}{2\sigma_y^2}\right)$$

1.3 多元高斯分布

$$p(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\ \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

其中，μ为各随机变量的均值组成的$n \times 1$向量，Σ表示随机变量间的$n \times n$协方差矩阵，正定。

二、多元高斯分布的条件概率密度

令随机向量$X = [x_1, x_2, \ldots, x_n]$服从多元高斯分布$X \sim N(\mu, \Sigma)$，令$X_1 = [x_1, \ldots, x_m]$为已经观测变量，$X_2 = [x_{m+1}, \ldots, x_n]$为未知变量，则：

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

从而有：

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$$\Sigma = \begin{bmatrix} \Sigma_{11}, \Sigma_{12} \\ \Sigma_{21}, \Sigma_{22} \end{bmatrix}$$

给定X_1求X_2的后验分布（这部分推导可以从相关文献中查到，此处略）：

$$\mu_{2|1} = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(X_1 - \mu_1)$$

$$\Sigma_{2|1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

三、高斯过程回归

设随机变量$X = [x_1, x_2, x_3, \ldots, x_n]^T$，其服从正态分布：

$$X \sim N(\mu, \Sigma)$$

其中$\mu = [\mu_0, \mu_1, \ldots, \mu_n]^T$为均值向量，$\Sigma$是这$n$个特征之间的协方差矩阵，将$\Sigma$展开有：

$$\Sigma = \begin{bmatrix} cov_{1,1}, & cov_{1,2}, \ldots, & cov_{1,n} \\ cov_{2,1}, & cov_{2,2}, \ldots, & cov_{2,n} \\ \ldots, & \ldots, & \ldots \\ cov_{n,1}, & cov_{n,2}, \ldots, & cov_{n,n} \end{bmatrix}$$

其中$cov_{i,j}$表示特征i和特征j之间的协方差（covariance）。

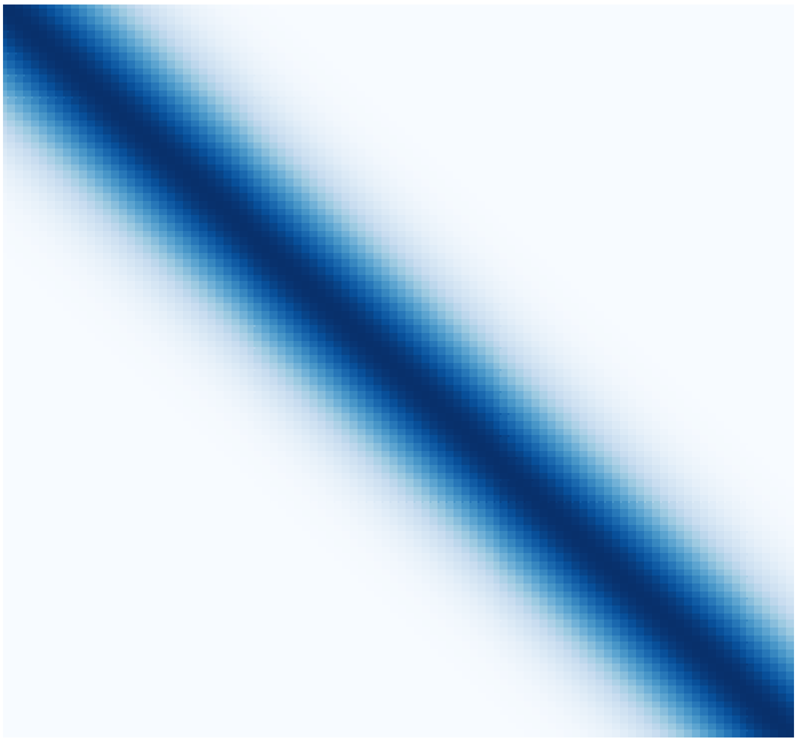
高斯过程样本与一般机器学习的样本区别在于，高斯过程中样本各特征之间存在相关关系，这种相关关系是通过协方差矩阵Σ来体现的。比如在一些时间序列模型里面，各个变量输出的时间序列在时间前后都会体现出一种相关性（比如平滑过渡等），这种模型输出就很适合使用高斯过程来模拟。

3.1 协方差矩阵计算

Σ可以通过高斯过程核进行求解，常见的高斯过程核有：

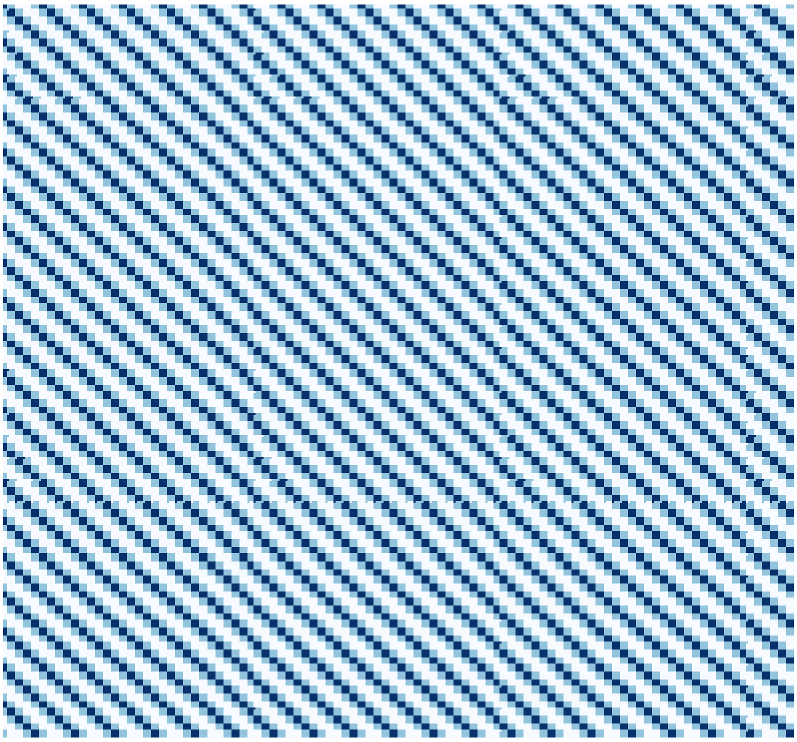
RBF kernel:

$$k = \sigma^2 \exp\left(-\frac{||t_a - t_b||^2}{2l^2}\right)$$



periodic kernel:

$$k = \sigma^2 \exp\left(-\frac{2}{l^2}\sin\left(\frac{\pi}{p}\right)|t_a - t_b|\right)$$



linear_kernel:

$$k = \sigma_b^2 + \sigma^2 * (t_a - c)(t_b - c)$$

这样当知道两个随机变量指数t_a和t_b后，便可通过核函数计算两个变量间的协方差。如果对所有随机变量均进行上述计算便可获得协方差矩阵Σ。有了协方差矩阵Σ后便可对高斯过程进行采样（一般认为高斯过程先验分布均值μ应无偏为0）。

3.2 高斯过程采样

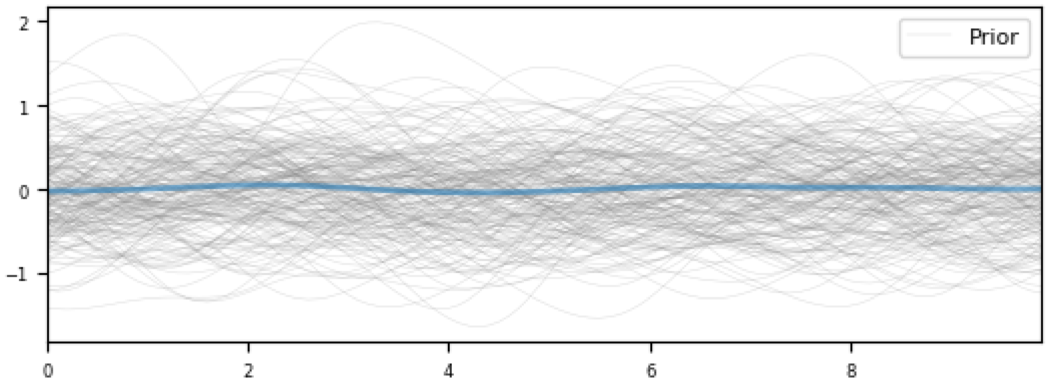
获得了各随机变量x的均值信息μ和联合分布的协方差矩阵Σ后，便可对该高斯过程进行随机采样。采样步骤如下：

1. 首先对协方差矩阵Σ进行SVD分解，获得矩阵U、S和V；
2. 生成N个独立同分布的高斯随机变量（均值为0，标准差为1），组成向量y；

3. 按照如下公式获得高斯过程样本：

$$x = \mu + \mathrm{U}\sqrt{\mathrm{S}}y$$

采样结果如下图所示，图中每条灰色曲线便对应一条高斯过程样本（$n=100$）,蓝色曲线表示样本均值，因为我们设定先验分布各维度上均值$\mu_i=0$，所以蓝色曲线在0附近波动。



3.3 后验分布和采样

3.2中获得的高斯过程样本为先验样本。但是当我们在某些指数t上获得了一批观测样本后，这批观测样本将有助于我们对其他指数集上的样本分布进行估计（后验）。我们将这批已观测指数集设为X_1，未观测到的指数集设为X_2。接下来便可使用第二节中的方法获得在X_2上样本分布后验概率参数$\mu_{2|1}$和$\Sigma_{2|1}$，最后重新对X_2上的随机变量进行采样。下图显示了后验分布样本：