

# 高斯过程回归

将一维高斯分布推广到多变量中就得到了高斯网络，将多变量推广到无限维，就得到了高斯过程，高斯过程是定义在连续域（时间空间）上的无限多个高维随机变量所组成的随机过程。

在时间轴上的任意一个点都满足高斯分布吗，将这些点的集合叫做高斯过程的一个样本。

对于时间轴上的序列  $\xi_t$ ，如果  $\forall n \in N^+, t_i \in T$ ，有  $\xi_{t_1-t_n} \sim \mathcal{N}(\mu_{t_1-t_n}, \Sigma_{t_1-t_n})$ ，那么  $\{\xi_t\}_{t \in T}$  是一个高斯过程。

高斯过程有两个参数（高斯过程存在性定理），均值函数  $m(t) = \mathbb{E}[\xi_t]$  和协方差函数  $k(s, t) = \mathbb{E}[(\xi_s - \mathbb{E}[\xi_s])(\xi_t - \mathbb{E}[\xi_t])]$ 。

我们将贝叶斯线性回归添加核技巧的这个模型叫做高斯过程回归，高斯过程回归分为两种视角：

1. 权空间的视角-核贝叶斯线性回归，相当于  $x$  为  $t$ ，在每个时刻的高斯分布来源于权重，根据上面的推导，预测的函数依然是高斯分布。
2. 函数空间的视角-高斯分布通过函数  $f(x)$  来体现。

## 核贝叶斯线性回归

贝叶斯线性回归可以通过加入核函数的方法来解决非线性函数的问题，将  $f(x) = x^T w$  这个函数变为  $f(x) = \phi(x)^T w$ （当然这个时候， $\Sigma_p$  也要变为更高维度的），变换到更高维的空间，有：

$$f(x^*) \sim \mathcal{N}(\phi(x^*)^T \sigma^{-2} A^{-1} \Phi^T Y, \phi(x^*)^T A^{-1} \phi(x^*)) \quad (1)$$

$$A = \sigma^{-2} \Phi^T \Phi + \Sigma_p^{-1} \quad (2)$$

其中， $\Phi = (\phi(x_1), \phi(x_2), \dots, \phi(x_N))^T$ 。

为了求解  $A^{-1}$ ，可以利用 Woodbury Formula， $A = \Sigma_p^{-1}, C = \sigma^{-2} \mathbb{I}$ ：

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1} \quad (3)$$

所以  $A^{-1} = \Sigma_p - \Sigma_p \Phi^T (\sigma^2 \mathbb{I} + \Phi \Sigma_p \Phi^T)^{-1} \Phi \Sigma_p$

也可以用另一种方法：

$$\begin{aligned} A &= \sigma^{-2} \Phi^T \Phi + \Sigma_p^{-1} \\ \Leftrightarrow A \Sigma_p &= \sigma^{-2} \Phi^T \Phi \Sigma_p + \mathbb{I} \\ \Leftrightarrow A \Sigma_p \Phi^T &= \sigma^{-2} \Phi^T \Phi \Sigma_p \Phi^T + \Phi^T = \sigma^{-2} \Phi^T (k + \sigma^2 \mathbb{I}) \\ \Leftrightarrow \Sigma_p \Phi^T &= \sigma^{-2} A^{-1} \Phi^T (k + \sigma^2 \mathbb{I}) \\ \Leftrightarrow \sigma^{-2} A^{-1} \Phi^T &= \Sigma_p \Phi^T (k + \sigma^2 \mathbb{I})^{-1} \\ \Leftrightarrow \phi(x^*)^T \sigma^{-2} A^{-1} \Phi^T &= \phi(x^*)^T \Sigma_p \Phi^T (k + \sigma^2 \mathbb{I})^{-1} \end{aligned} \quad (4)$$

上面的左边的式子就是变换后的均值，而右边的式子就是不含  $A^{-1}$  的式子，其中  $k = \Phi \Sigma_p \Phi^T$ 。

根据  $A^{-1}$  得到方差为：

$$\phi(x^*)^T \Sigma_p \phi(x^*) - \phi(x^*)^T \Sigma_p \Phi^T (\sigma^2 \mathbb{I} + k)^{-1} \Phi \Sigma_p \phi(x^*) \quad (5)$$

上面定义了：

$$k = \Phi \Sigma_p \Phi^T \quad (6)$$

我们看到，在均值和方差中，含有下面四项：

$$\phi(x^*)^T \Sigma_p \Phi^T, \phi(x^*)^T \Sigma_p \phi(x^*), \phi(x^*)^T \Sigma_p \Phi^T, \Phi \Sigma_p \phi(x^*) \quad (7)$$

展开后，可以看到，有共同的项： $k(x, x') = \phi(x)^T \Sigma_p \phi(x')$ 。由于  $\Sigma_p$  是正定对称的方差矩阵，所以，这是一个核函数。

对于高斯过程中的协方差：

$$k(t, s) = \text{Cov}[f(x), f(x')] = \mathbb{E}[\phi(x)^T w w^T \phi(x')] = \phi(x)^T \mathbb{E}[w w^T] \phi(x') = \phi(x)^T \Sigma_p \phi(x') \quad (8)$$

我们可以看到，这个就对应着上面的核函数。因此我们看到  $\{f(x)\}$  组成的组合就是一个高斯过程。

## 函数空间的观点

相比权重空间，我们也可以直接关注  $f$  这个空间，对于预测任务，这就是类似于求：

$$p(y^* | X, Y, x^*) = \int_f p(y^* | f, X, Y, x^*) p(f | X, Y, x^*) df \quad (9)$$

对于数据集来说，取  $f(X) \sim \mathcal{N}(\mu(X), k(X, X))$ ,  $Y = f(X) + \varepsilon \sim \mathcal{N}(\mu(X), k(X, X) + \sigma^2 \mathbb{I})$ 。预测任务的目的是给定一个新数据序列  $X^* = (x_1^*, \dots, x_M^*)^T$ ，得到  $Y^* = f(X^*) + \varepsilon$ 。我们可以写出：

$$\begin{pmatrix} Y \\ f(X^*) \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \mu(X) \\ \mu(X^*) \end{pmatrix}, \begin{pmatrix} k(X, X) + \sigma^2 \mathbb{I} & k(X, X^*) \\ k(X^*, X) & k(X^*, X^*) \end{pmatrix} \right) \quad (10)$$

根据高斯分布的方法：

$$x = \begin{pmatrix} x_a \\ x_b \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}, \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \right) \quad (11)$$

$$x_b | x_a \sim \mathcal{N}(\mu_{b|a}, \Sigma_{b|a}) \quad (12)$$

$$\mu_{b|a} = \Sigma_{ba} \Sigma_{aa}^{-1} (x_a - \mu_a) + \mu_b \quad (13)$$

$$\Sigma_{b|a} = \Sigma_{bb} - \Sigma_{ba} \Sigma_{aa}^{-1} \Sigma_{ab} \quad (14)$$

可以直接写出：

$$\begin{aligned} p(f(X^*) | X, Y, X^*) &= p(f(X^*) | Y) \\ &= \mathcal{N}(k(X^*, X) [k(X, X) + \sigma^2 \mathbb{I}]^{-1} (Y - \mu(X)) + \mu(X^*), \\ &\quad k(X^*, X^*) - k(X^*, X) [k(X, X) + \sigma^2 \mathbb{I}]^{-1} k(X, X^*)) \end{aligned} \quad (15)$$

所以对于  $Y = f(X^*) + \varepsilon$ ：

$$\begin{aligned} &\mathcal{N}(k(X^*, X) [k(X, X) + \sigma^2 \mathbb{I}]^{-1} (Y - \mu(X)) + \mu(X^*), \\ &k(X^*, X^*) - k(X^*, X) [k(X, X) + \sigma^2 \mathbb{I}]^{-1} k(X, X^*) + \sigma^2 \mathbb{I} \end{aligned} \quad (17)$$

我们看到，函数空间的观点更加简单易于求解。