

9

Gaussian Process

高斯过程

用于机器学习中的非线性回归和分类



人类拥有海量史籍，但是不能操纵历史；

人类可能主宰未来，却对未来一无所知。

We may have knowledge of the past but cannot control it; we may control the future but have no knowledge of it.

—— 克劳德·香农 (Claude Shannon) | 美国数学家、工程师、密码学家 | 1916 ~ 2001



- ◀ `sklearn.gaussian_process.GaussianProcessRegressor()` 高斯过程回归函数
- ◀ `sklearn.gaussian_process.kernels.RBF()` 高斯过程高斯核函数
- ◀ `sklearn.gaussian_process.GaussianProcessClassifier()` 高斯过程分类函数

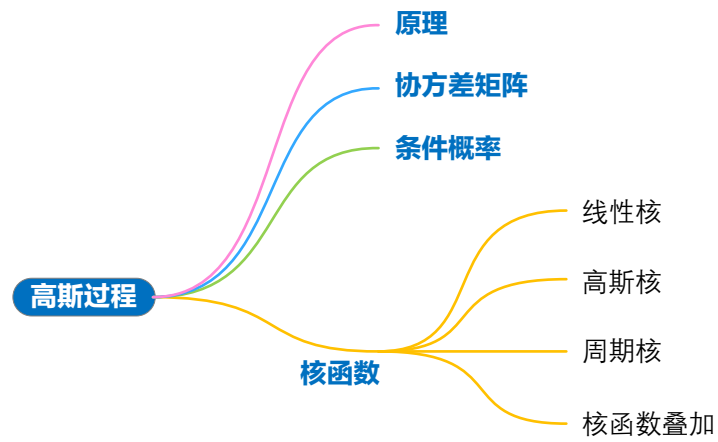
本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com



9.1 高斯过程原理

高斯过程 (Gaussian Process, GP) 是一种概率模型，用于建模连续函数或实数值变量的概率分布。在高斯过程中，任意一组数据点都可以被视为多元高斯分布的样本，该分布的均值和协方差矩阵由先验信息和数据点间的相似度计算而得。通过高斯过程，可以对函数进行预测并对其不确定性进行量化，这使得其在机器学习、优化和贝叶斯推断等领域中被广泛应用。

在使用高斯过程进行预测时，通常使用条件高斯分布来表示先验和后验分布。通过先验分布和数据点的观测，可以计算后验分布，并通过该分布来预测新数据点的值。在高斯过程中，协方差函数或核函数起着重要的作用，它定义了数据点间的相似性，不同的核函数也适用于不同的应用场景。一些常见的核函数包括线性核、多项式核、高斯核、拉普拉斯核等。

本章将首先以**高斯核** (Gaussian kernel) 为例，介绍如何理解高斯过程算法原理。注意，高斯核也叫**径向基核** (Radial Basis Function kernel, RBF kernel)。

先验

\mathbf{x}_2 为一系列需要预测的点， $\mathbf{y}_2 = \text{GP}(\mathbf{x}_2)$ 对应高斯过程预测结果。

高斯过程的先验为：

$$\mathbf{y}_2 \sim N(\boldsymbol{\mu}_2, \mathbf{K}_{22}) \quad (1)$$

其中， $\boldsymbol{\mu}_2$ 为高斯过程的均值（通常默认为全 0 向量）， \mathbf{K}_{22} 为协方差矩阵。之所以写成 \mathbf{K}_{22} 这种形式，是因为高斯过程的协方差矩阵通过核函数定义。

在 Scikit-learn 中，高斯核的定义为：

$$\kappa(x_i, x_j) = \text{cov}(y_i, y_j) = \exp\left(-\frac{(x_i - x_j)^2}{2l^2}\right) \quad (2)$$

图 1 所示为 $l = 1$ 时先验协方差矩阵的热图。

为了保证形式上和协方差矩阵一致，图 1 纵轴上下调转。这个协方差矩阵显然是对称矩阵，它的主对角线是**方差** (variance)，非主对角线元素为**协方差** (covariance)。

回到 (2)，不难发现 $\kappa(x_i, x_j)$ 体现的是 y_i 和 y_j 的协方差（即描述协同运动），但是 $\kappa(x_i, x_j)$ 通过 x_i 和 x_j 两个坐标点确定的。更确切地说，如图 2 所示，当 l 一定时， x_i 和 x_j 间距绝对值 ($\Delta x = x_i - x_j$) 越大， $\kappa(x_i, x_j)$ 越小；反之， Δx 越小， $\kappa(x_i, x_j)$ 越大。这一点后续将会反复提及。

此外，图 2 还展示了参数 l 对高斯函数的影响。

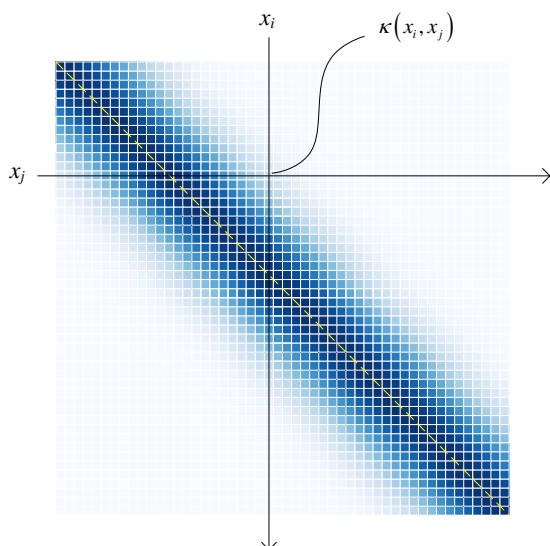


图 1. 高斯过程的先验协方差矩阵，高斯核

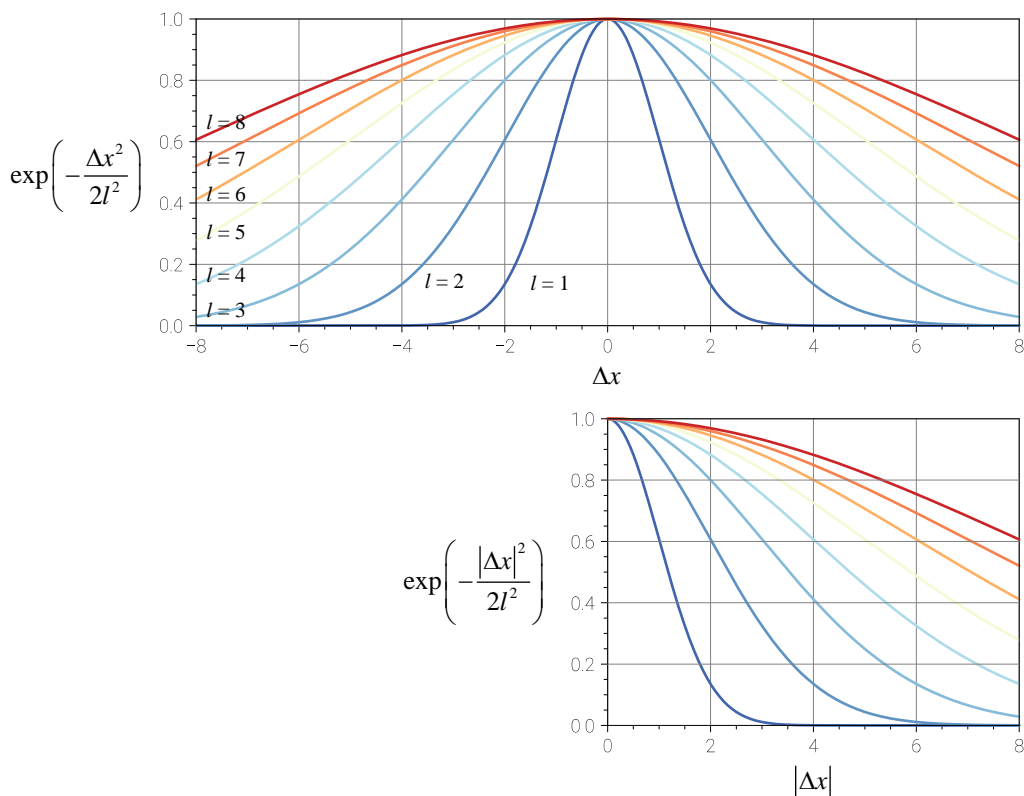
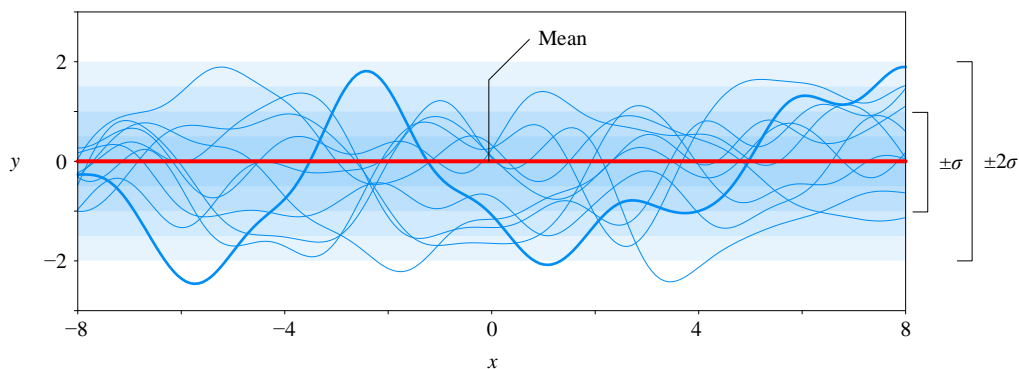
图 2. 高斯核函数受参数 l 的影响

图 3 所示每一条代表一个根据当前先验均值、先验协方差的函数采样。打个比方，在没有引入数据之前，图 3 的曲线可以看成是一捆没有扎紧的丝带，随着微风飘动。

图 3 中的红线为高斯过程的先验均值，本章假设均值为 0。本章接下来要解释为什么图 3 中曲线是这种形式。

图 3. 高斯过程的采样，高斯核先验协方差矩阵， $\sigma = 1$

样本数据

观测到的样本数据为 (x_1, y_1) 。图 4 给出 5 个样本点，大家很快就会发现这 5 个点相当于扎紧丝带的 5 个节点。下面，我们要用贝叶斯方法来帮我们整合“先验 + 数据”，并计算后验分布。

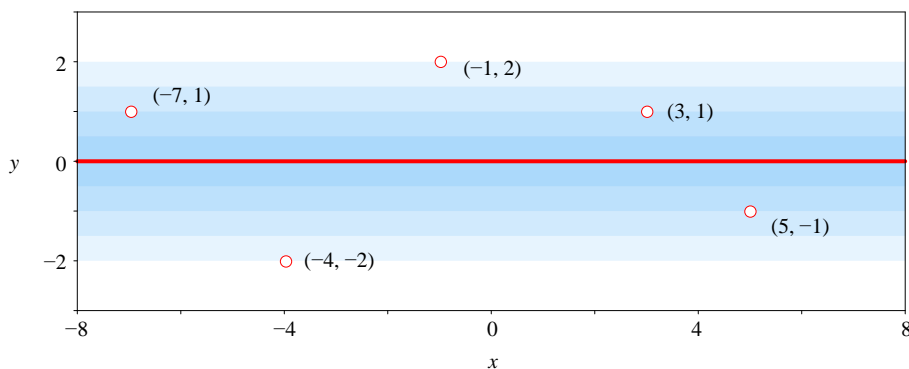


图 4. 给定 5 个样本数据

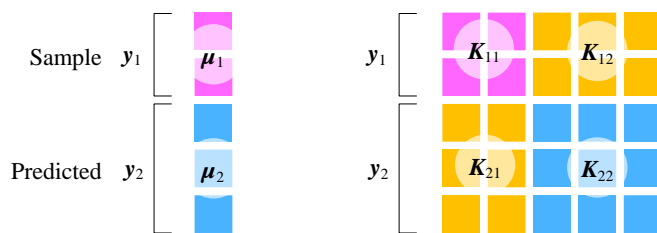
联合分布

假设样本数据 y_1 和预测值 y_2 服从联合高斯分布：

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}\right) \quad (3)$$

简单来说，高斯过程对应的分布可以看成是无限多个随机变量的联合分布。图 5 中的协方差矩阵来自 $[x_1, x_2]$ 的核函数。本章后文会用实例具体展示如何计算上式中的协方差矩阵 (K_{11}, K_{22}) 和互协方差矩阵 (K_{12}, K_{21}) 。

⚠ 注意，一般假设 μ_1 、 μ_2 为全 0 向量。

图 5. 样本数据 y_1 和预测值 y_2 服从联合高斯分布

后验分布

根据条件高斯分布，我们可以获得后验分布为：

$$f(y_2 | y_1) \sim N \left(\underbrace{K_{21} K_{11}^{-1} (y_1 - \mu_1) + \mu_2}_{\text{Expectation}}, \underbrace{K_{22} - K_{21} K_{11}^{-1} K_{12}}_{\text{Covariance matrix}} \right) \quad (4)$$

看到这个式子，特别是条件期望部分，大家是否想到了多元线性回归？

如图 6 所示，在 5 个样本点位置丝带被锁紧，而其余部分丝带仍然舞动。

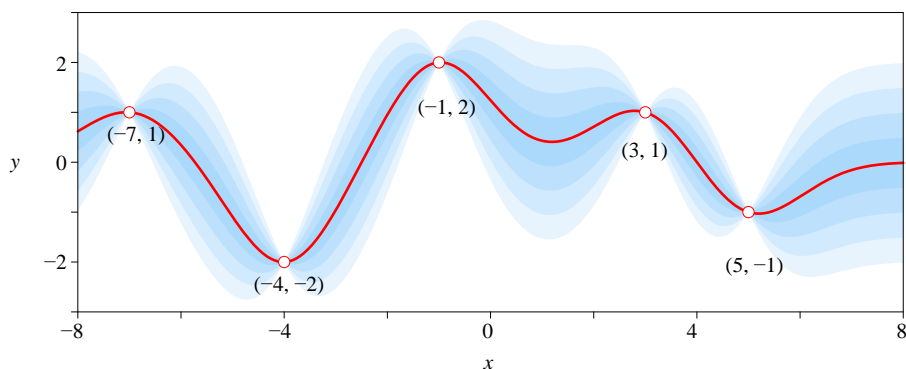


图 6. 高斯过程后验分布的采样函数，高斯核

图 6 中红色曲线对应后验分布的均值：

$$K_{21} K_{11}^{-1} (y_1 - \mu_1) + \mu_2 \quad (5)$$

图 6 中带宽对应一系列标准差：

$$\text{sqrt}(\text{diag}(K_{22} - K_{21} K_{11}^{-1} K_{12})) \quad (6)$$

其中， $\text{diag}()$ 表示获取对角线元素； $\text{sqrt}()$ 代表开平方得到一组标准差序列，代表纵轴位置的不确定性。

如图 7 所示，在高斯过程算法中，贝叶斯定理将先验和数据整合到一起得到后验。看到这里，大家如果还是不理解高斯过程原理，不要紧。下面，我们就用这个例子展开讲解高斯过程中的技术细节。

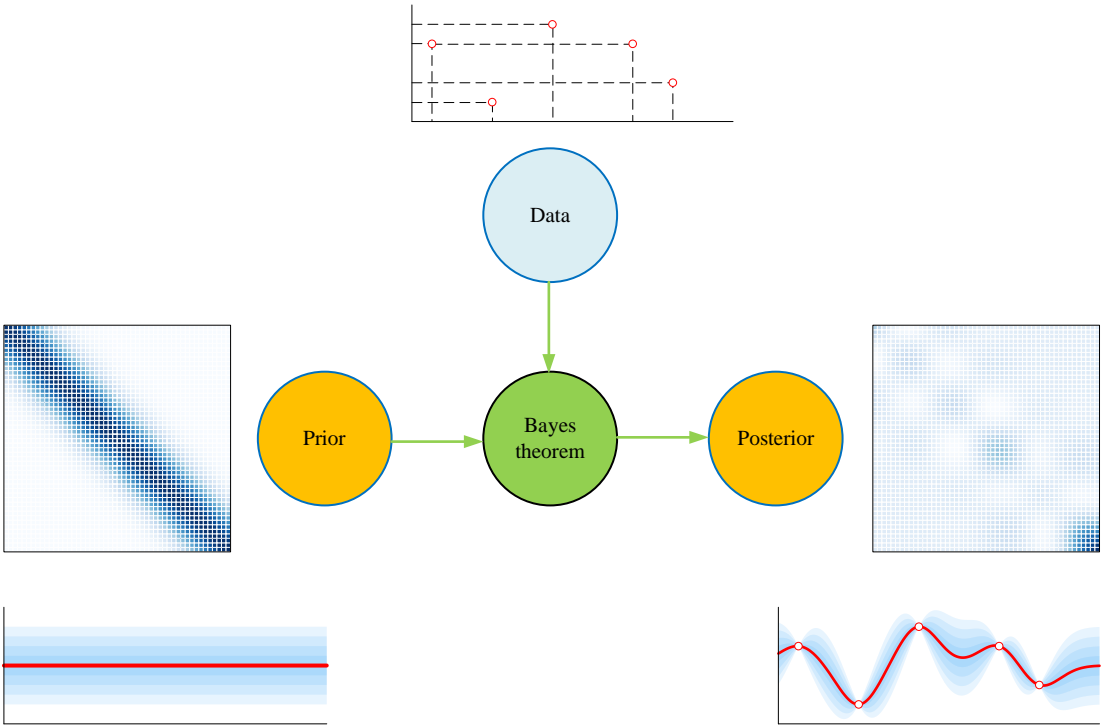


图 7. 高斯过程算法中贝叶斯定理的作用

9.2 协方差矩阵

基于高斯核的协方差矩阵相当是一种“人造”协方差矩阵。当然，这种“人造”协方差矩阵有它的独到之处，下面就近距离观察这个协方差矩阵。

以往，我们看到的协方差矩阵都是有限大小，通常用热图表示。高斯过程算法用到的协方差矩阵实际上是无限大。比如，如果 x 的取值范围为 $[-8, 8]$ ，在这个区间内满足条件的 x 值有无数个。

如图 8 所示，我们用三维网格图呈现这个无限大的协方差矩阵。和一般的协方差矩阵一致，这个协方差矩阵的主对角线元素为方差，非主对角线元素为协方差。

我们容易发现图 8 这个协方差矩阵特别像是一个二元函数。我们可以固定一个变量，看协方差值随另外一个变量变化。不难发现，图中的每条曲线都是一条高斯函数。

再次强调， x 为预测点，我们关注的是高斯过程预测结果 $y = GP(x)$ 之间的关系。简单来说， x 提供位置坐标，不同 $y = GP(x)$ 之间的协同运动用高斯核来描述。

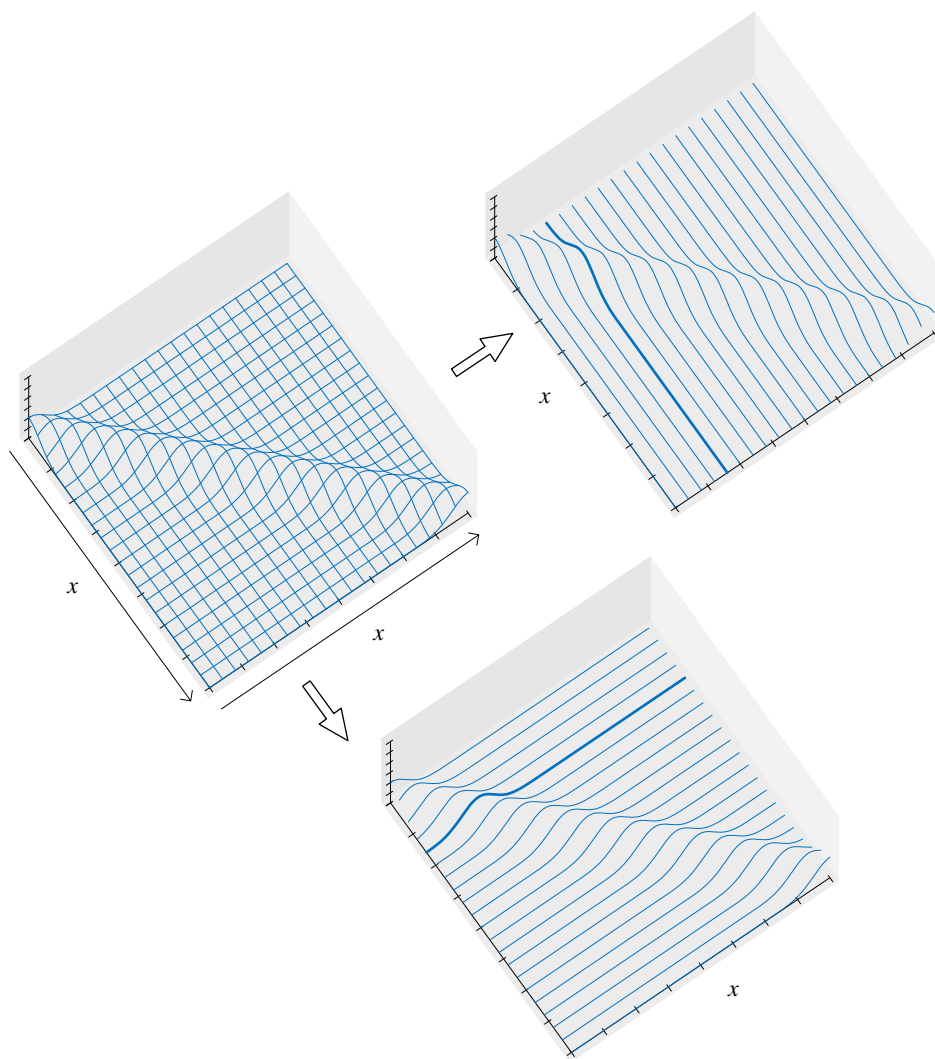


图 8. 无限大的先验协方差矩阵，高斯核

为了方便可视化，同时为了和我们熟悉的协方差矩阵对照来看，我们选取 $[-8, 8]$ 区间中 50 个点，并绘制如图 9 所示的协方差矩阵热图。下面，我们来观察图 9 中协方差矩阵的每一行。

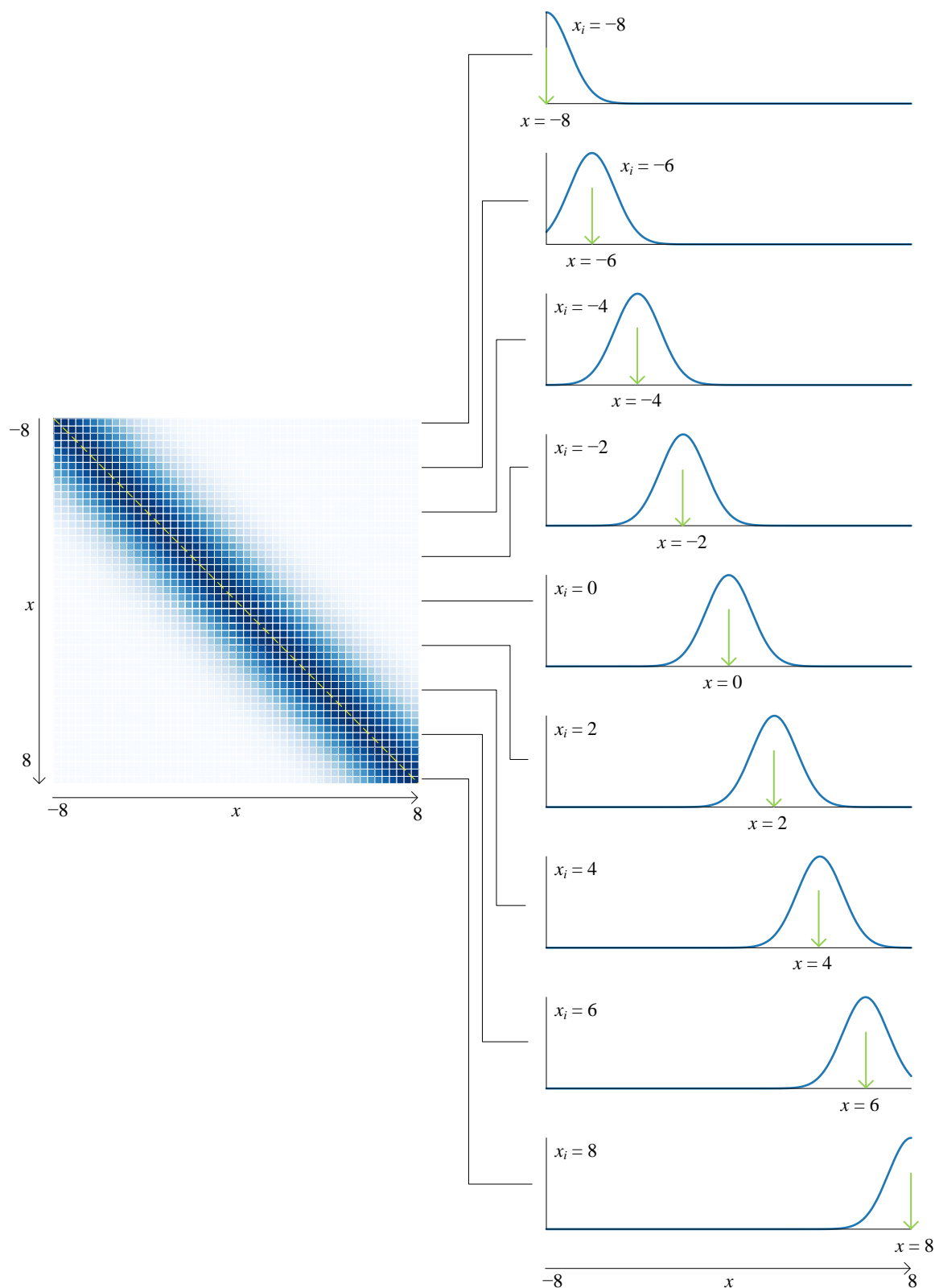


图 9. 观察协方差矩阵的每一行

当确定一个 x_i 取值后，比如 $x_i = 0$ ，对于区间 $[-8, 8]$ 上任意一点 x ，利用高斯核我们都可以计算得到一个协方差值

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

$$\kappa(x_i, x) = \exp\left(-\frac{(x_i - x)^2}{2l^2}\right) \quad (7)$$

观察这个函数，我们可以发现，函数在 $x = x_i$ 取得最大值。而随着 x 不断远离 x_i ，即 $x_i - x$ 绝对值增大，协方差值不断减小，不断靠近 0。

对于 (7)， $x = x_i$ 这一点又恰好是 x_i 位置处的 $y_i = \text{GP}(x_i)$ 方差，即 $\text{var}(y_i) = \kappa(x_i, x_i) = 1$ 。有了这一观察，我们可以发现图 10 中协方差主对角线元素都是 1；也就是说，在给定均值为 0，高斯核为先验函数的条件下，任何一点处 x_i 的 $y_i = \text{GP}(x_i)$ 具有相同的“不确定性”。这就是我们可以在图 3 中观察到的，“丝带”任何一点在一定范围内飘动。图 3 的每条“丝带”上的纵轴值服从多元高斯分布。

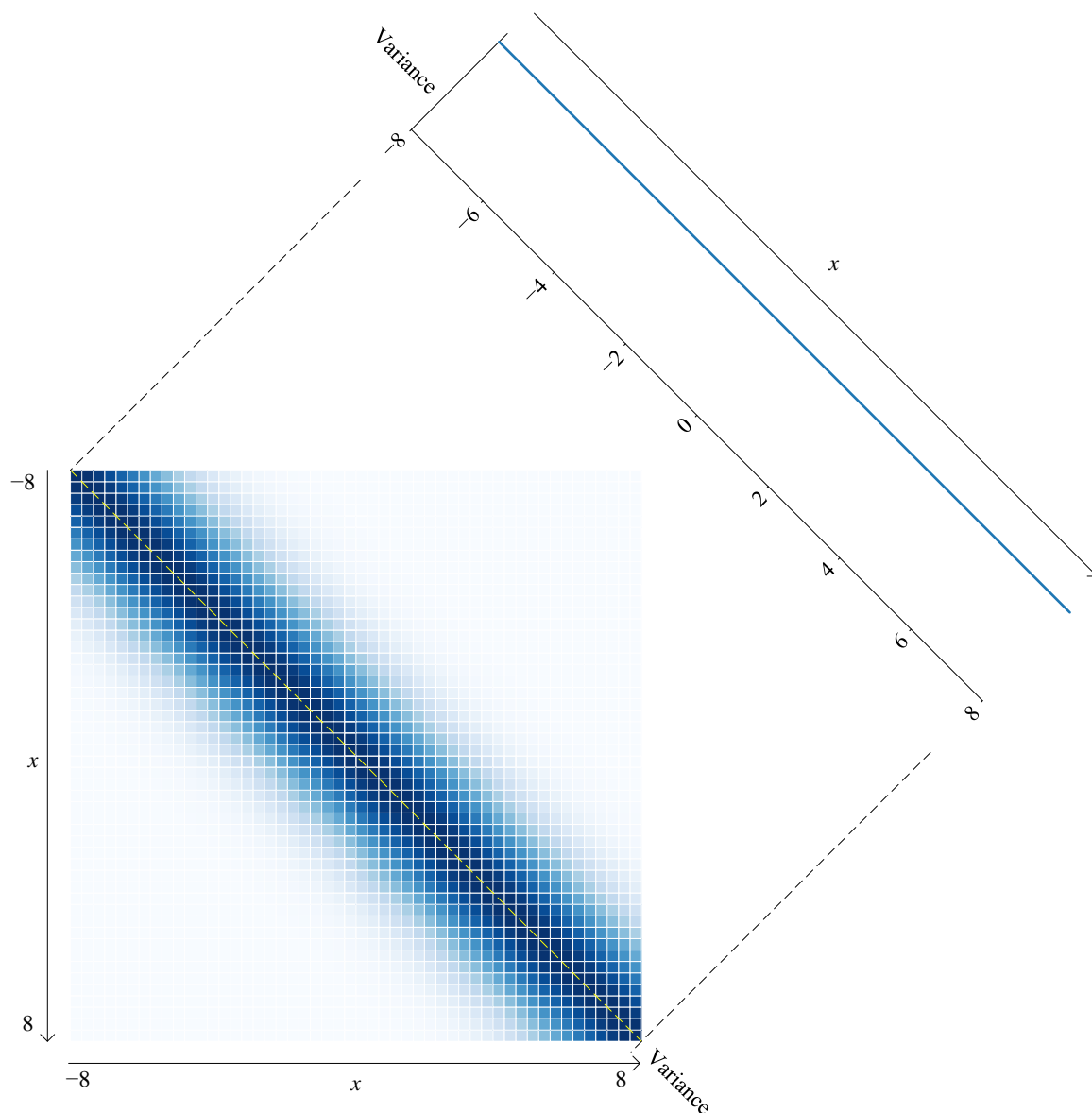


图 10. 高斯核协方差矩阵的方差 (对角元素)

但是观察图 3，我们还发现同一条丝带看上去很“顺滑”，这又是为什么？

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

想要理解这一点，我们就要关注协方差矩阵中的协方差成分。

在给定 (2) 这种形式的高斯核条件下，对于点 x_i ，它和 $x_i + \Delta x$ 的 2×2 协方差矩阵可以写成。

$$\begin{bmatrix} 1 & \exp\left(-\frac{\Delta x^2}{2l^2}\right) \\ \exp\left(-\frac{\Delta x^2}{2l^2}\right) & 1 \end{bmatrix} \quad (8)$$

我们已经知道，点 x_i 和点 $x_i + \Delta x$ 的方差都是 1，因此两者的相关性系数为 $\exp\left(-\frac{\Delta x^2}{2l^2}\right)$ 。

当 l 为定值时， Δx 绝对值越大，即点 $x_i + \Delta x$ 离 x_i 越远，两者的相关性越靠近 0；相反， Δx 绝对值越小，即点 $x_i + \Delta x$ 离 x_i 越近，两者的相关性越靠近 1。

如图 11 所示，当相关性系数 $\exp\left(-\frac{\Delta x^2}{2l^2}\right)$ (非负值) 为不同值时，代表 2×2 协方差矩阵 (8) 的椭圆不断变化。

这就解释了图 3 中每一条丝带看上去很顺滑的原因。越靠近丝带的任意一点，相关性越高，也就是说具有更高的协同运动；距离特定点越远，相关性越低，协同运动关系也就越差。

请大家注意，不限定取值范围时， x_i 可以是实轴上任意一点；换个角度来看， x_i 有无数个。

很多其他文献上中高斯核定义为：

$$\kappa(x_i, x_j) = \sigma^2 \exp\left(-\frac{(x_i - x_j)^2}{2l^2}\right) \quad (9)$$

上式中先验协方差矩阵中的方差不再是 1，而是 σ^2 。

大家可能会问既然这个高斯核协方差矩阵是“人造”的，我们可不可以创造其他形式的协方差矩阵？

答案是肯定的！本章最后将介绍高斯过程中常用的其他核函数。

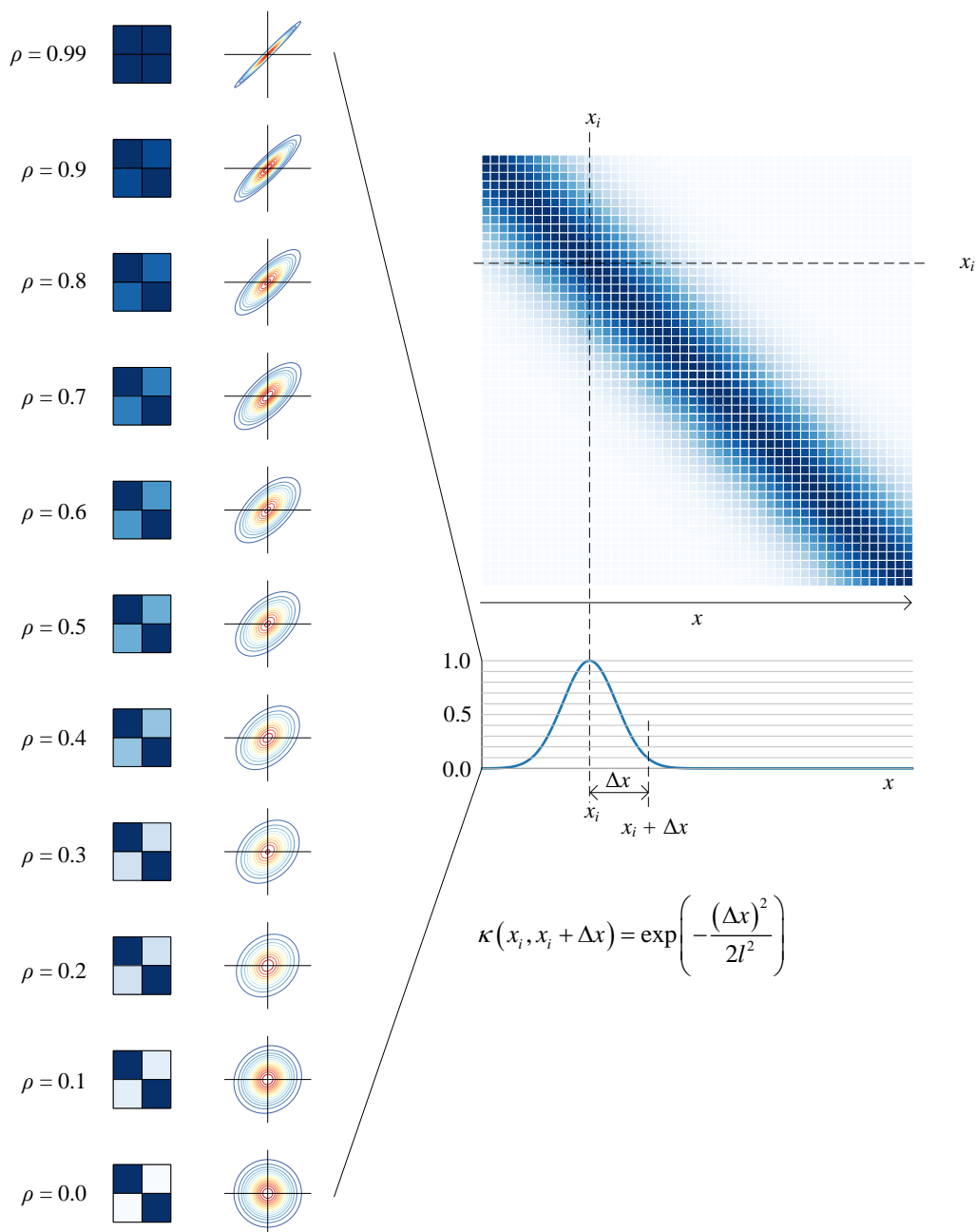


图 11. 高斯核协方差矩阵的协方差 (非对角元素)

9.3 分块协方差矩阵

根据 (3)，我们先将协方差矩阵分块，具体如图 12 所示。其中， \mathbf{K}_{11} 是样本数据的协方差矩阵， \mathbf{K}_{22} 是先验协方差矩阵。 \mathbf{K}_{12} 和 \mathbf{K}_{21} 都是互协方差矩阵 (cross covariance matrix)，且互为转置。

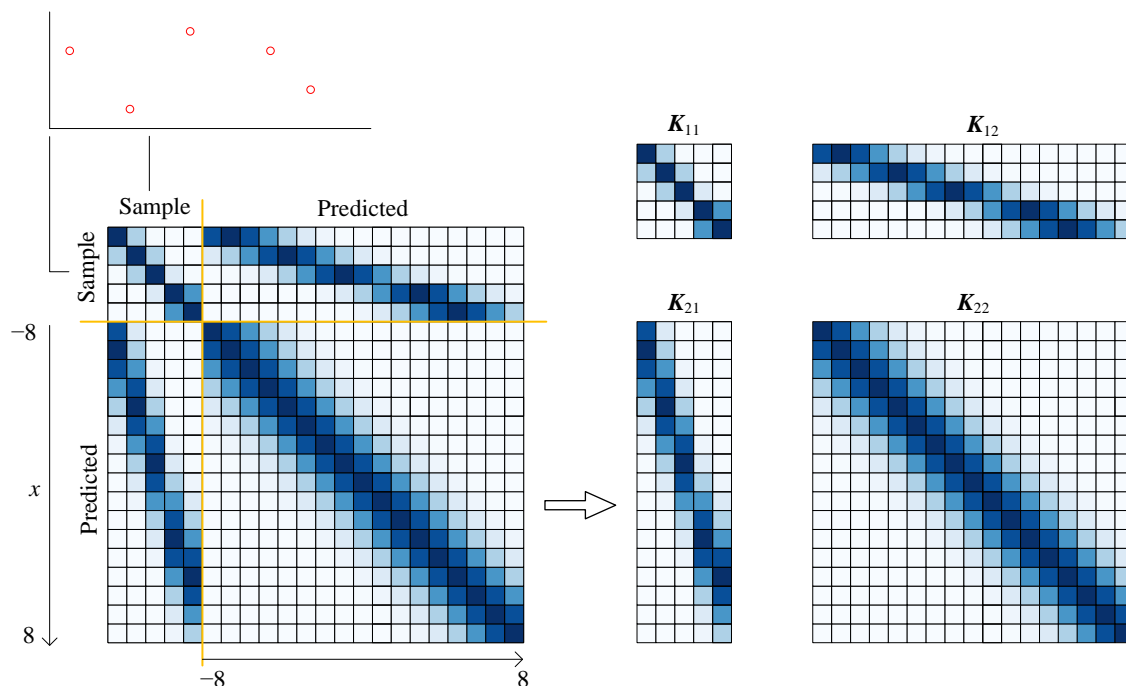
本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

图 12. 协方差矩阵 \mathbf{K} 分块

9.4 后验

下面就是利用 (4)，计算条件期望向量和条件协方差矩阵。

图 13 所示为计算条件期望向量 $\mathbf{K}_{21}\mathbf{K}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1) + \boldsymbol{\mu}_2$ 的过程。

\mathbf{y}_1 向量对应图 4 中 5 个红色点的 y 值序列。

默认， $\boldsymbol{\mu}_1$ 向量为全 0 向量，和 \mathbf{y}_1 形状相同。

图 13 中协方差矩阵 \mathbf{K}_{11} 则图 4 中 5 个红色点的 x 序列 (\mathbf{x}_1) 采用 (2) 计算得到。图 14 所示为计算 \mathbf{K}_{11} 的示意图。再次强调 $(\mathbf{x}_1, \mathbf{y}_1)$ 代表样本数据。

\mathbf{K}_{21} 的行代表 $[-8, 8]$ 区间上顺序采样的一组数值，即预测点序列 \mathbf{x}_2 ； \mathbf{K}_{21} 的列代表 5 个红色点的 x 值序列。图 15 所示为计算互协方差矩阵 \mathbf{K}_{21} 的示意图。显然， \mathbf{K}_{21} 是根据 \mathbf{x}_1 和 \mathbf{x}_2 计算得到的。

计算结果 $\mathbf{K}_{21}\mathbf{K}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1) + \boldsymbol{\mu}_2$ 为列向量，对应中图 6 中红色线 y 值序列。

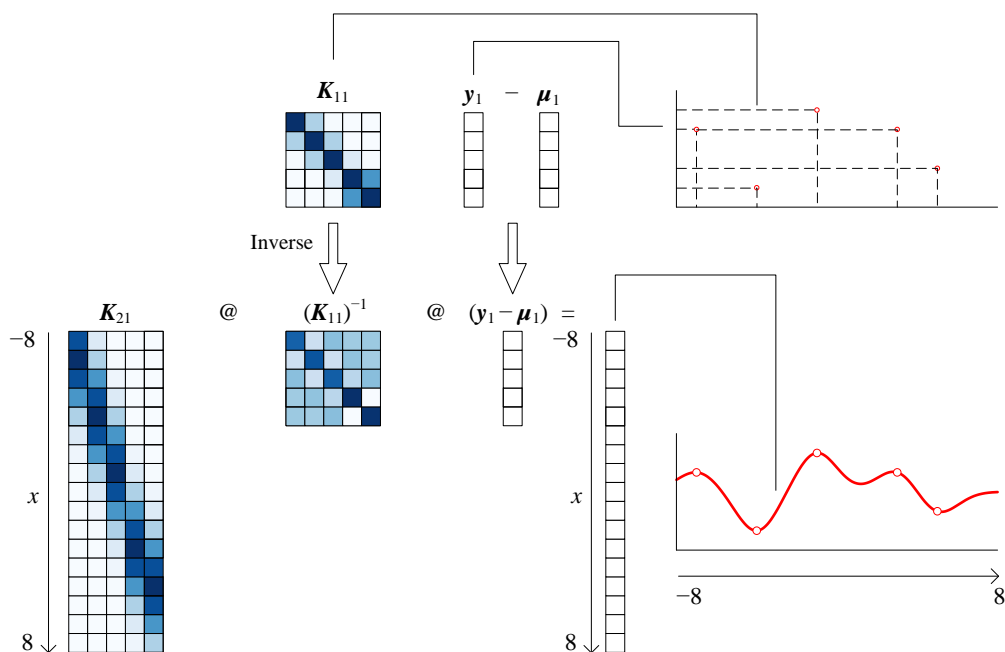


图 13. 计算条件期望

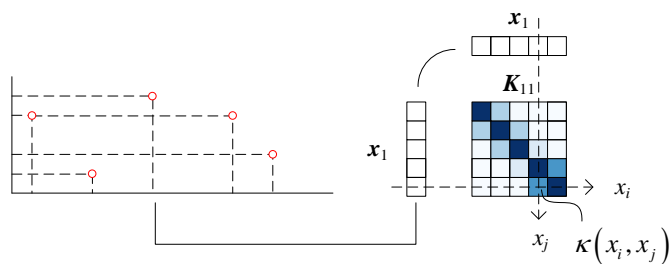
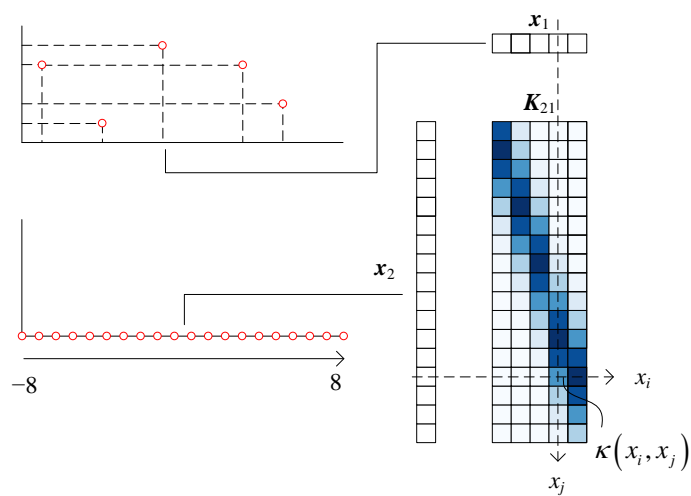
图 14. 计算协方差矩阵 K_{11} 图 15. 计算互协方差矩阵 K_{21}

图 16 所示为计算条件协方差 $K_{22} - K_{21}K_{11}^{-1}K_{12}$ 的过程。比较图 15 和图 17，很容易发现 K_{21} 和 K_{12} 互为转置。图 18 所示为自己算协方差矩阵的示意图 K_{22} 。

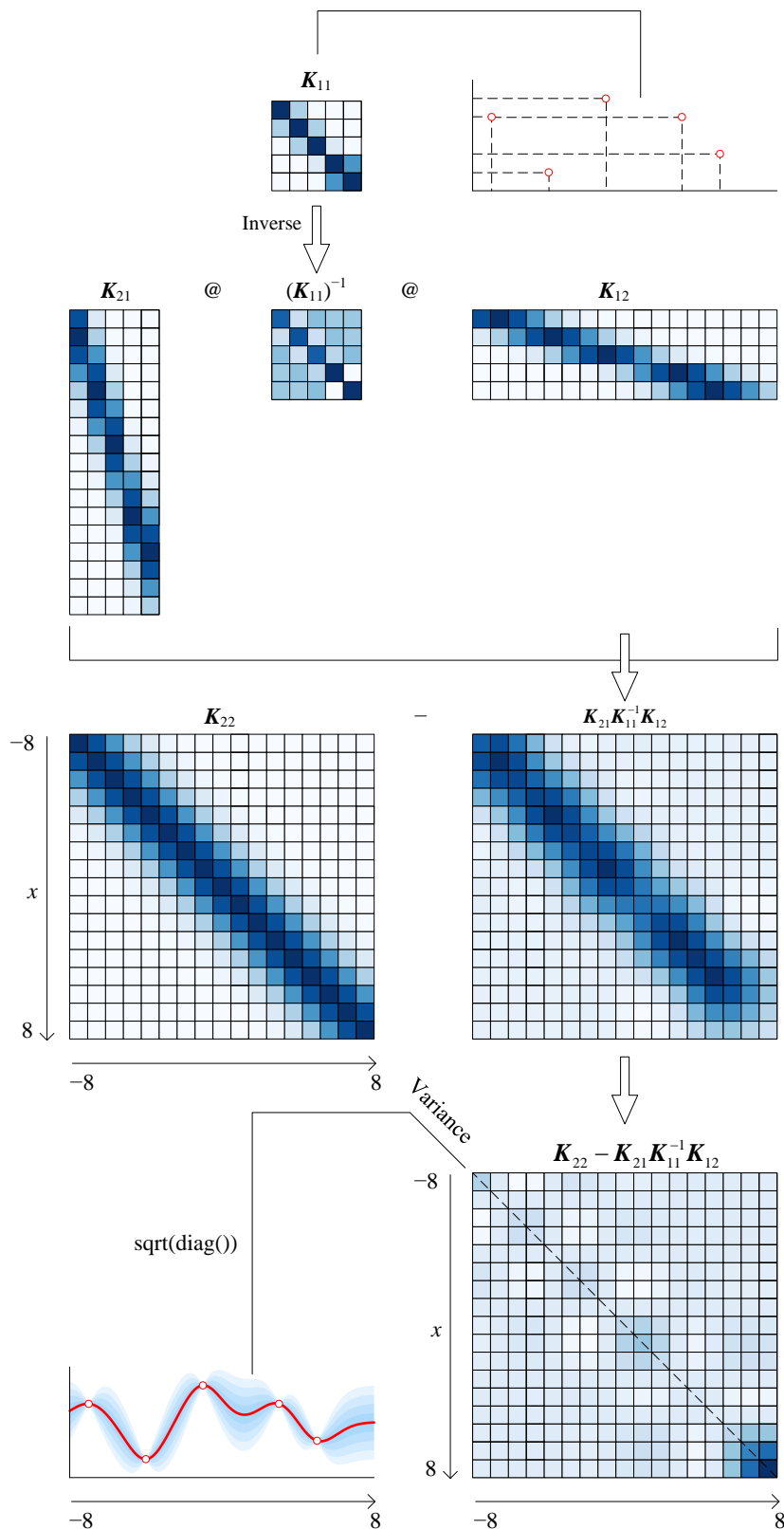


图 16. 计算条件协方差矩阵

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger: <https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

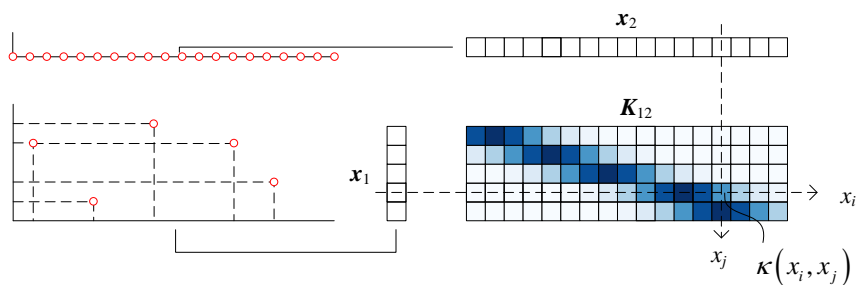
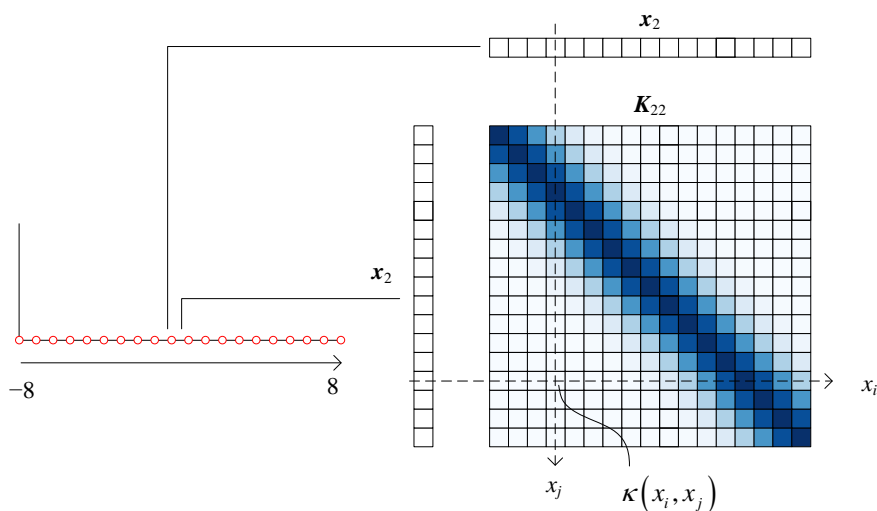
图 17. 计算互协方差矩阵 K_{12} 图 18. 计算协方差矩阵 K_{22}

图 19 所示为“无限大”的后验协方差矩阵曲面。和图 8 中先验协方差矩阵相比，我们可以发现在存在样本数据的位置，曲面发生了明显的塌陷。特别是在对角线上。

为了更清楚地看到这一点，我们特别绘制了图 20。后验协方差矩阵 $K_{22} - K_{21}K_{11}^{-1}K_{12}$ 的对角线元素为后验方差，体现了 x_2 不同位置上 y_2 值的不确定性。

图 20 中的后验标准差在 5 个数据点位置下降到 0；也就是说，不确定性为 0。这就解释了图 6 中 5 个“扎紧”的节点。

把图 6 和图 20 中的后验标准差放在同一张图上，我们便得到图 21。图 21 更方便展示上述现象。

换个角度来看，这也说明模型样本数据不存在任何“噪音”。倘若“噪音”存在，我们就需要修正 K_{11} 。这是下一节要介绍的内容。

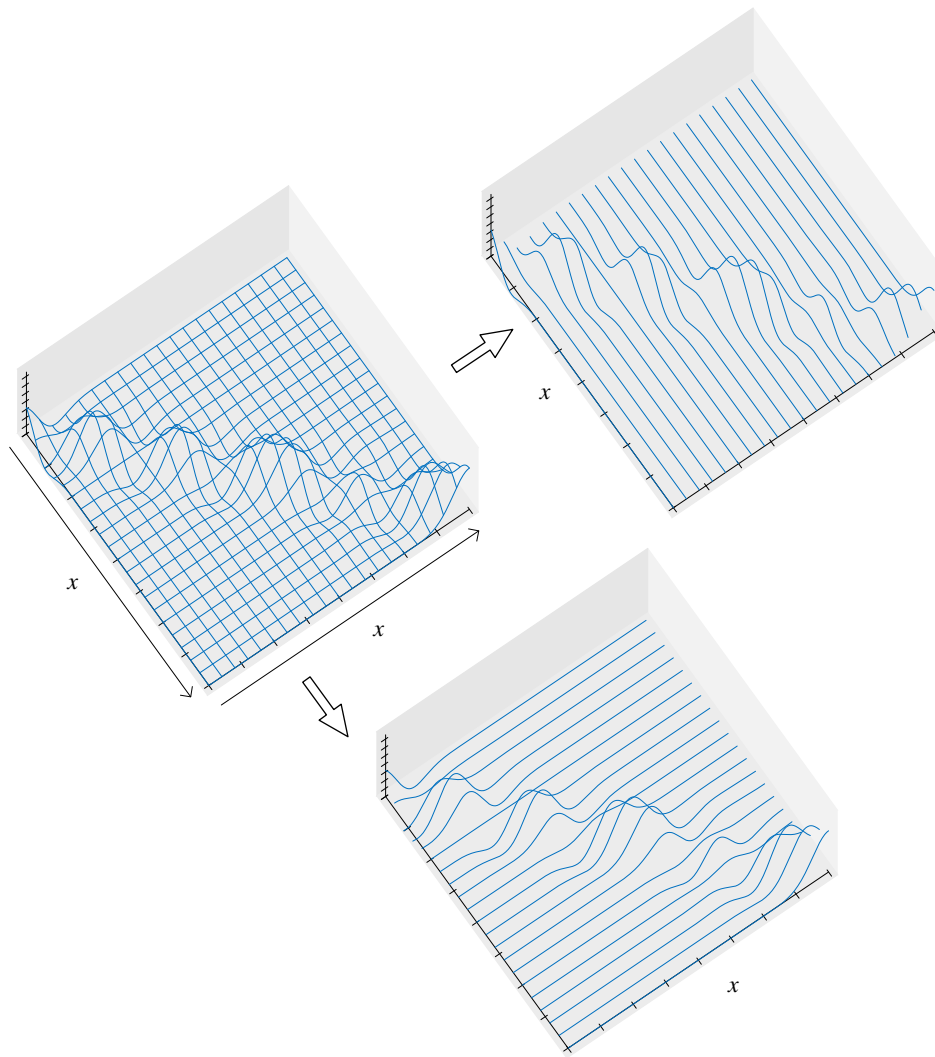


图 19. 无限大的后验协方差矩阵

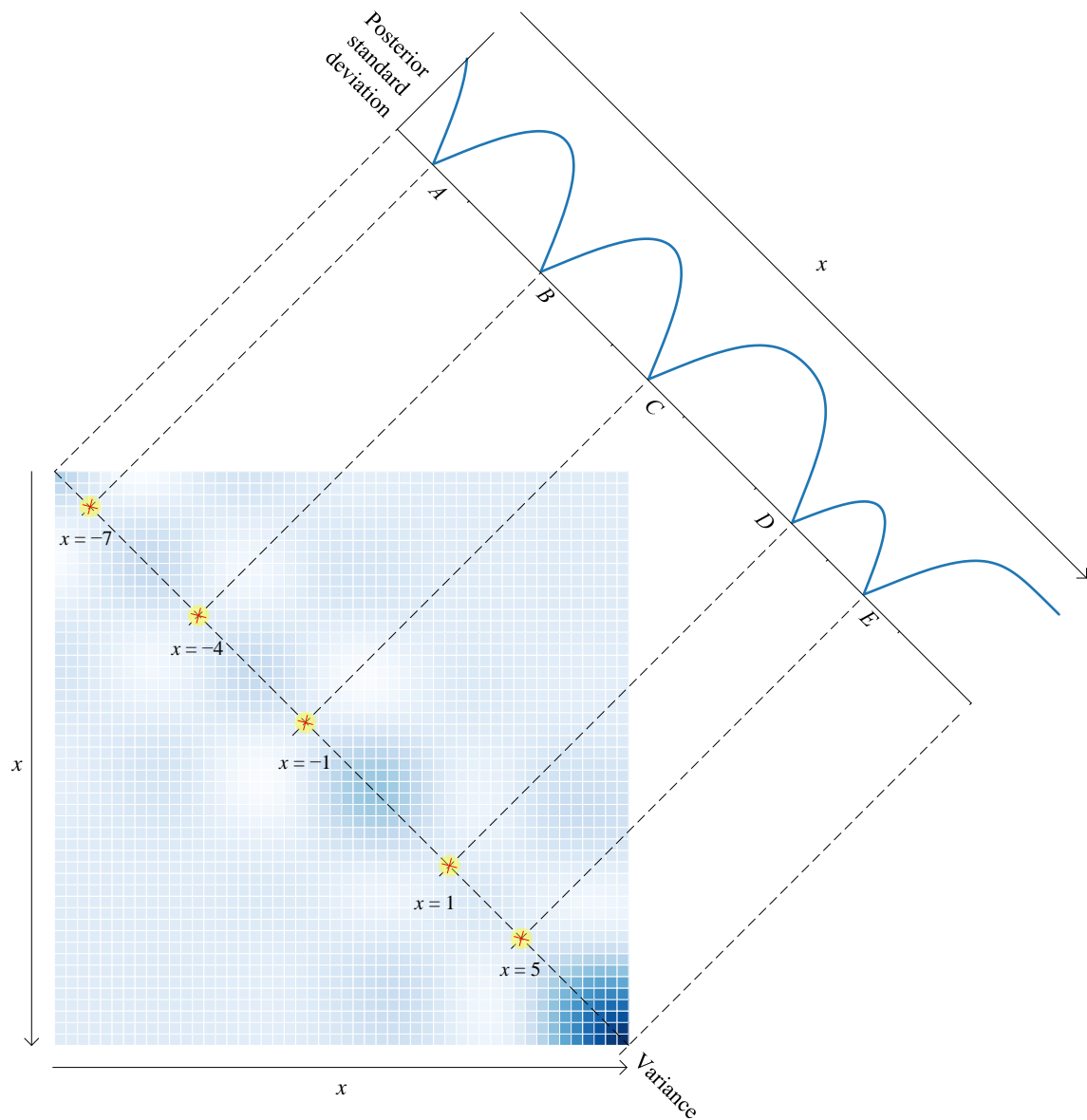


图 20. 后验协方差矩阵对角线元素

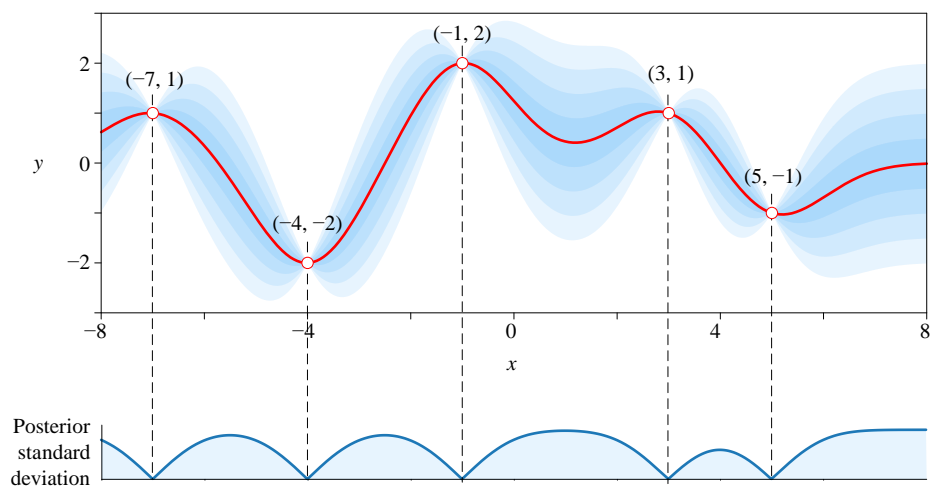


图 21. 高斯过程后验分布

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

9.5 噪音

上一节提到图 20 中给定数据点处后验方差将至 0，这意味着数据不存在噪音。反之，当说数据存在噪音时，意味着观测到的数据可能受到随机误差或不确定性的影响。

在高斯过程中处理带有噪音的数据通常包括在模型中的 \mathbf{K}_{11} 中引入噪声项 (如图 22)，以反映实际观测中的不确定性。图 23 所示为不同噪音水平对结果影响。

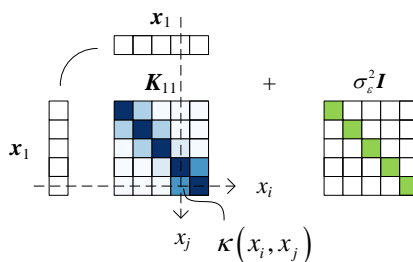


图 22. 在协方差矩阵 \mathbf{K}_{11} 加噪音

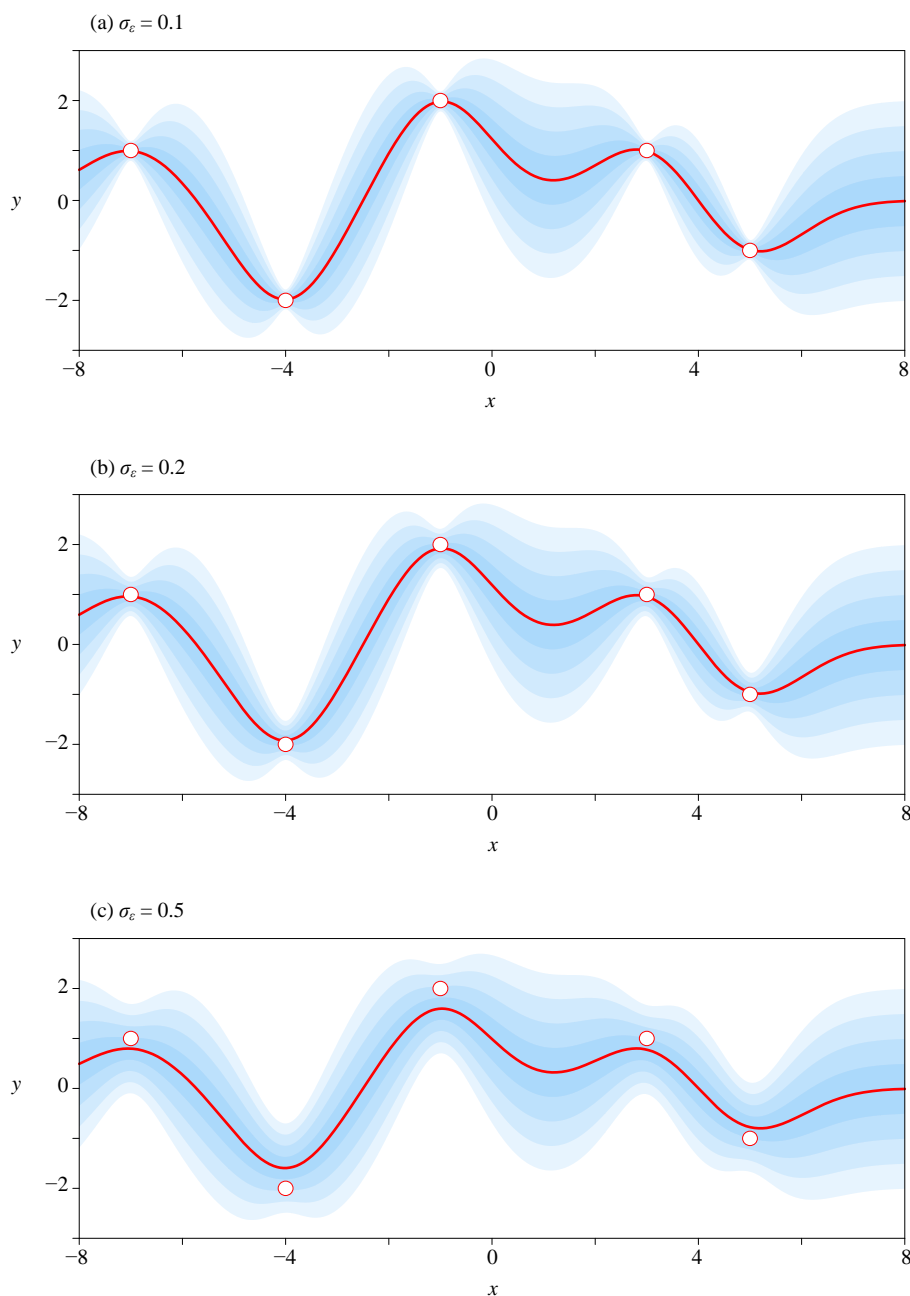


图 23. 在噪音对结果影响

9.6 核函数

核函数 (kernel function) 是机器学习中一个常用概念。我们在**高斯过程** (Gaussian Process, GP)、**支持向量机** (Support Vector Machine, SVM)、**核主成分分析** (Kernel Principal Component Analysis, KPCA) 中用到核函数。本节则侧重高斯过程中的核函数。《机器学习》将专门介绍支持向量机中的**核技巧** (kernel trick)，以及核主成分分析中的核函数。

经过本章前文的学习，大家已经清楚高斯过程是一种用于回归和分类的非参数模型，它通过对输入空间中数据点之间的相似性进行建模，从而实现输出的推断。

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger: <https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

核函数在高斯过程中的作用是定义输入空间中数据点之间的相似性或相关性。以前文介绍过的高斯核为例，核函数决定了在输出空间中，两个数据点对应的预测值之间的协方差。

如图 1 所示，如果两个数据点 (x_i, x_j) 在输入空间中相似，它们对应的输出值就有更高的协方差（深蓝色），反之则有较低的协方差（浅蓝色）。注意，上述描述适用于高斯核协方差矩阵，并不能描述其他常见核函数。表 1 总结高斯过程中常见的核函数。

下面，我们展开聊聊线性核、高斯核、周期核。本节最后还会介绍核函数叠加。

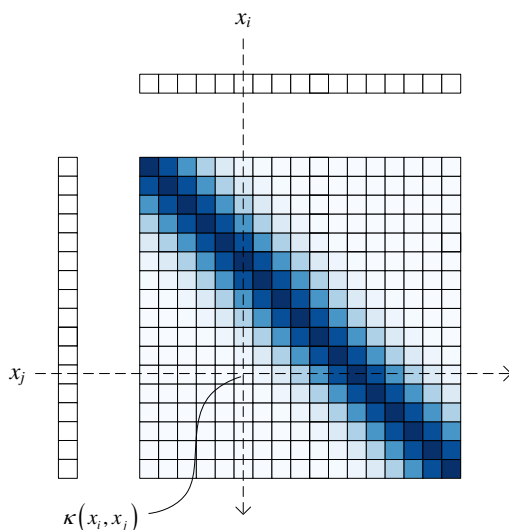
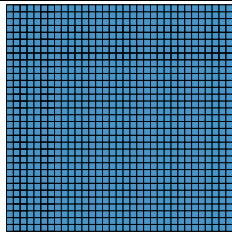
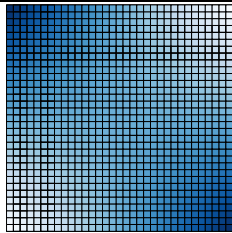
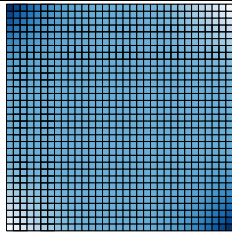
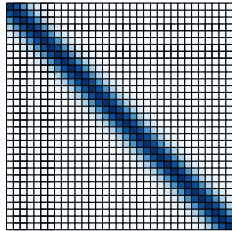
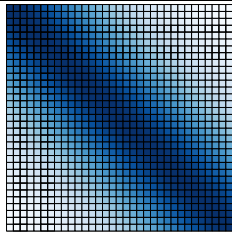
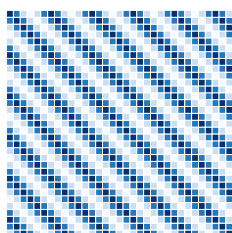


图 24. 高斯过程的高斯核协方差矩阵

表 1. 高斯过程中常用的核函数

核函数	常见形式	先验协方差矩阵
常数核 (constant kernel)	$\kappa(x_i, x_j) = c$	
线性核 (linear kernel)	$\kappa(x_i, x_j) = \sigma^2 (x_i - c)(x_j - c)$ $\kappa(x_i, x_j) = \sigma_b^2 + \sigma^2 (x_i - c)(x_j - c)$	

多项式核 (polynomial kernel)	$\kappa(x_i, x_j) = ((x_i - c)(x_j - c) + \text{offset})^d$ $\kappa(x_i, x_j) = (\sigma^2 (x_i - c)(x_j - c) + \text{offset})^d$	
高斯核 (Gaussian kernel)	$\kappa(x_i, x_j) = \exp\left(-\frac{(x_i - x_j)^2}{2l^2}\right)$ $\kappa(x_i, x_j) = \sigma^2 \exp\left(-\frac{(x_i - x_j)^2}{2l^2}\right)$	
有理二次核 (rational quadratic kernel)	$\kappa(x_i, x_j) = \exp\left(1 + \frac{(x_i - x_j)^2}{2\alpha l^2}\right)^{-\alpha}$ $\kappa(x_i, x_j) = \sigma^2 \exp\left(1 + \frac{(x_i - x_j)^2}{2\alpha l^2}\right)^{-\alpha}$	
周期核 (periodic kernel)	$\kappa(x_i, x_j) = \sigma^2 \exp\left(-\frac{2 \sin^2\left(\frac{\pi}{p} x_i - x_j \right)}{l^2}\right)$ $\kappa(x_i, x_j) = \sigma^2 \exp\left(-\frac{\sin^2\left(\frac{\pi}{p} x_i - x_j \right)}{2l^2}\right)$	

线性核

线性核是高斯过程中的一种核函数，也称为线性相似性函数。它是一种用于衡量输入数据点之间线性关系的核函数。本节采用的线性核函数的表达式为

$$\kappa(x_i, x_j) = \text{cov}(y_i, y_j) = \sigma^2 (x_i - c)(x_j - c) \quad (10)$$

其中， x_i 和 x_j 分别是输入空间中的两个数据点。图 8 所示为线性核先验协方差矩阵的曲面。

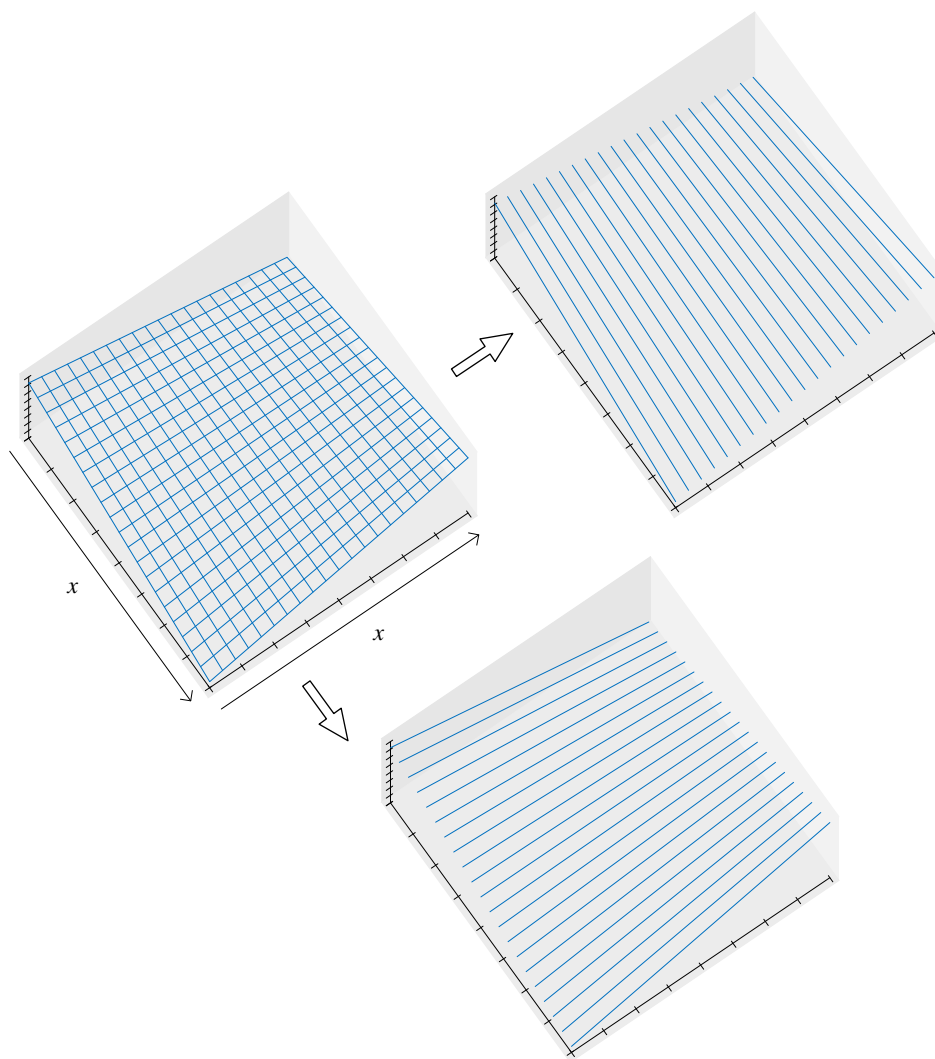
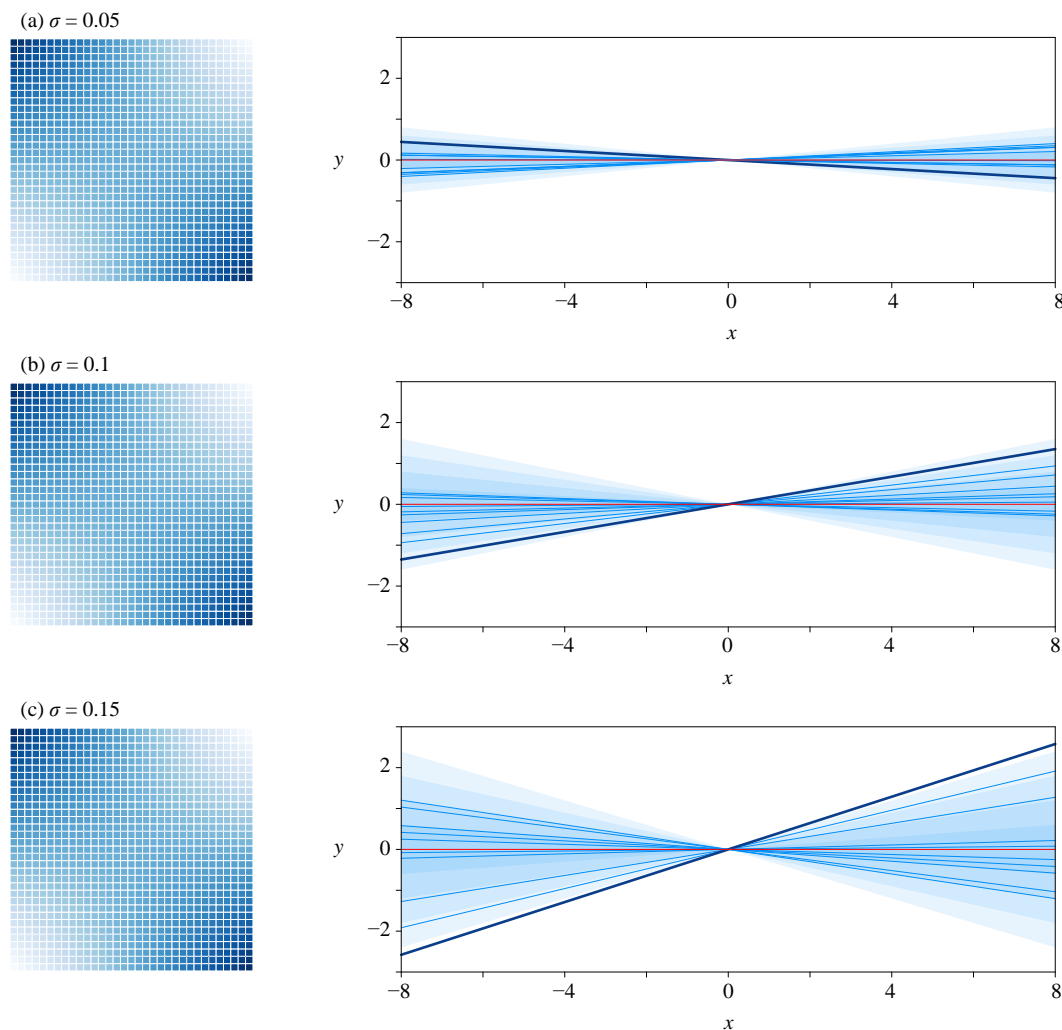


图 25. 无限大的先验协方差矩阵，线性核

图 26 所示为当 $c = 0$ 时，线性核参数 σ 对先验协方差和采样影响。注意，图 26 中不同热图子图的颜色映射取值范围不同。

线性核函数假设输入空间中的数据点之间存在线性关系，即输出值之间的相似性与输入数据点的线性组合有关。如果两个数据点在输入空间中更接近形成线性关系，它们对应的输出值在高斯过程中的协方差较高。

线性核函数在某些问题中很有效，特别是当数据呈现线性关系时。然而，对于非线性关系的数据，其他核函数如高斯核函数可能更适用，因为它们能够处理更复杂的数据结构。

图 26. 线性核参数 σ 对先验协方差和采样影响

高斯核

高斯核是在高斯过程中常用的核函数之一，也称为**径向基函数** (Radial Basis Function, RBF) 或**指数二次核** (exponentiated quadratic kernel, squared exponential)。本节采用的高斯核的形式为

$$\kappa(x_i, x_j) = \text{cov}(y_i, y_j) = \sigma^2 \exp\left(-\frac{(x_i - x_j)^2}{2l^2}\right) \quad (11)$$

其中， σ^2 为协方差矩阵的方差； l 是高斯核长度尺度度量参数。请大家翻阅前文查看高斯核的先验协方差曲面。

前文提过，高斯核函数的作用是衡量两个输入点之间的相似性，当两点距离较近时，核函数的值较大，表示它们在函数空间中具有相似的输出；反之，距离较远时核函数的值较小，表示它们在输出上差异较大。

图 27 所示为当 $\sigma = 1$ 时高斯核参数 l 对先验协方差和采样影响。比较几幅子图，我们可以发现 l 越大，协方差矩阵中协方差相对更大（临近点协同运动越强），对应曲线越平滑。

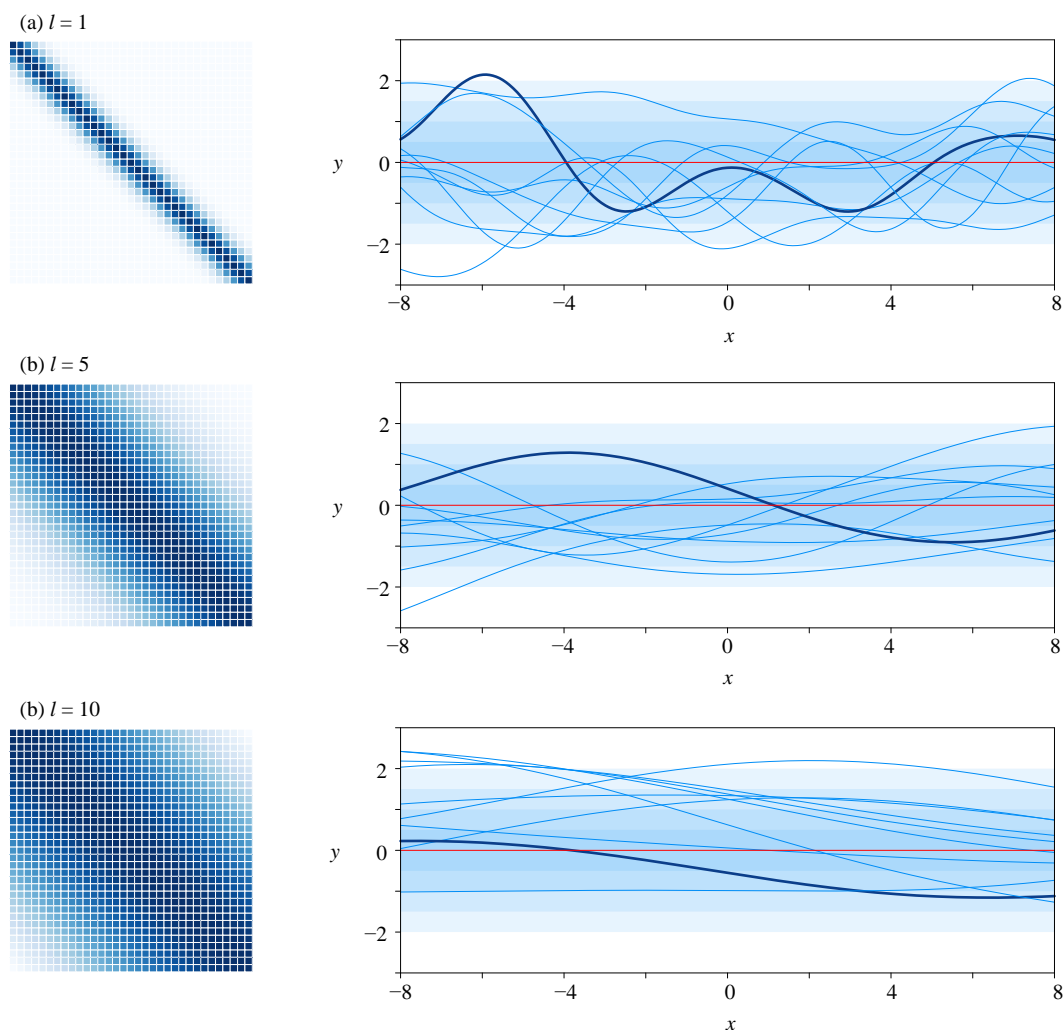
本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

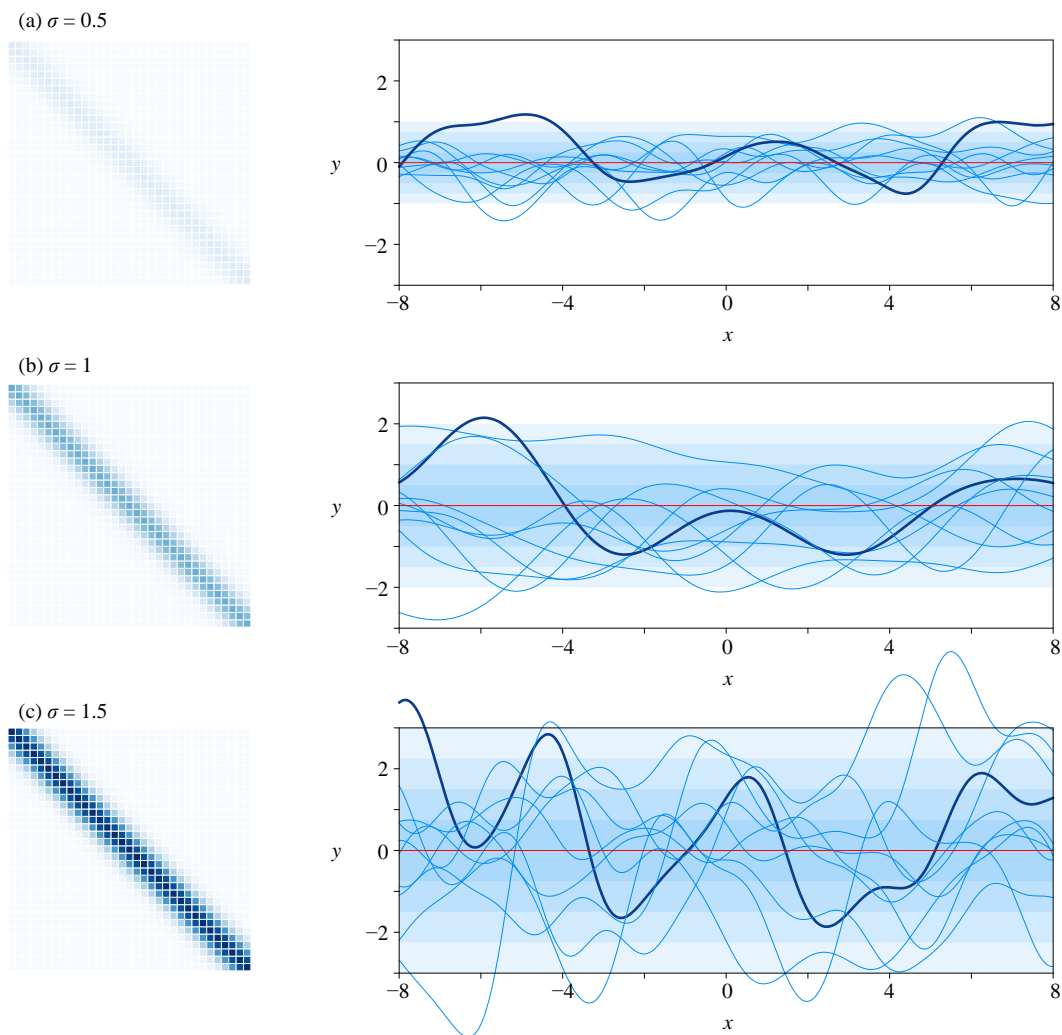
本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

图 27. 高斯核参数 l 对先验协方差和采样影响, $\sigma = 1$

对于 (11), 我们发现 σ 也是高斯核参数之一。图 28 所示为 $l = 1$ 时高斯核参数 σ 对先验协方差和采样影响。

高斯核具有平滑性和无限可微性的特点, 使其在建模各种复杂函数时表现出色。在高斯过程中, 选择合适的核函数和参数是关键, 它直接影响了模型对数据的拟合程度和泛化能力。

图 28. 高斯核参数 σ 对先验协方差和采样影响, $l = 1$

周期核

周期核是高斯过程中常用的核函数之一，它适用于描述具有周期性变化的数据。本节采用的周期核的形式为

$$\kappa(x_i, x_j) = \text{cov}(y_i, y_j) = \sigma^2 \exp \left(-\frac{\sin^2 \left(\frac{\pi}{p} |x_i - x_j| \right)}{2l^2} \right) \quad (12)$$

其中, p 影响周期核的周期, l 是高斯核的长度尺度参数。参数 l 对周期核的影响类似高斯核。

图 29 所示为周期核先验协方差曲面。

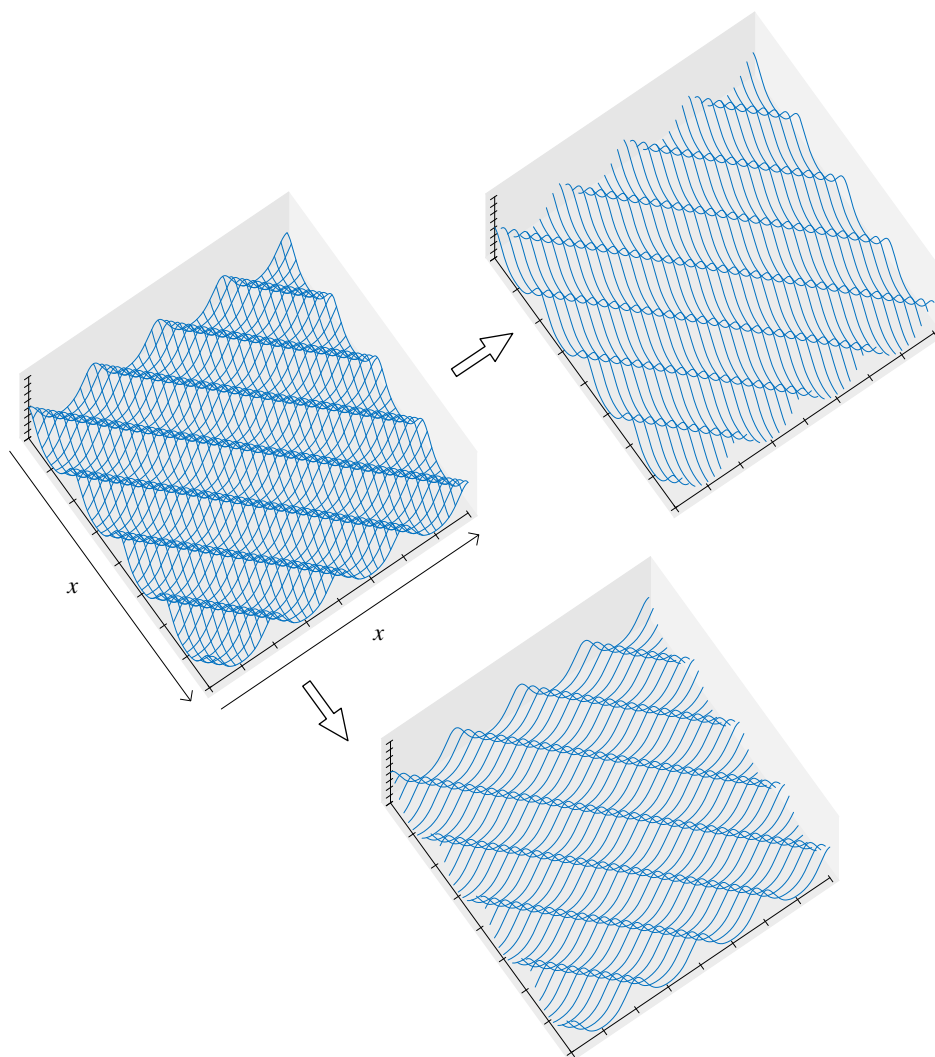


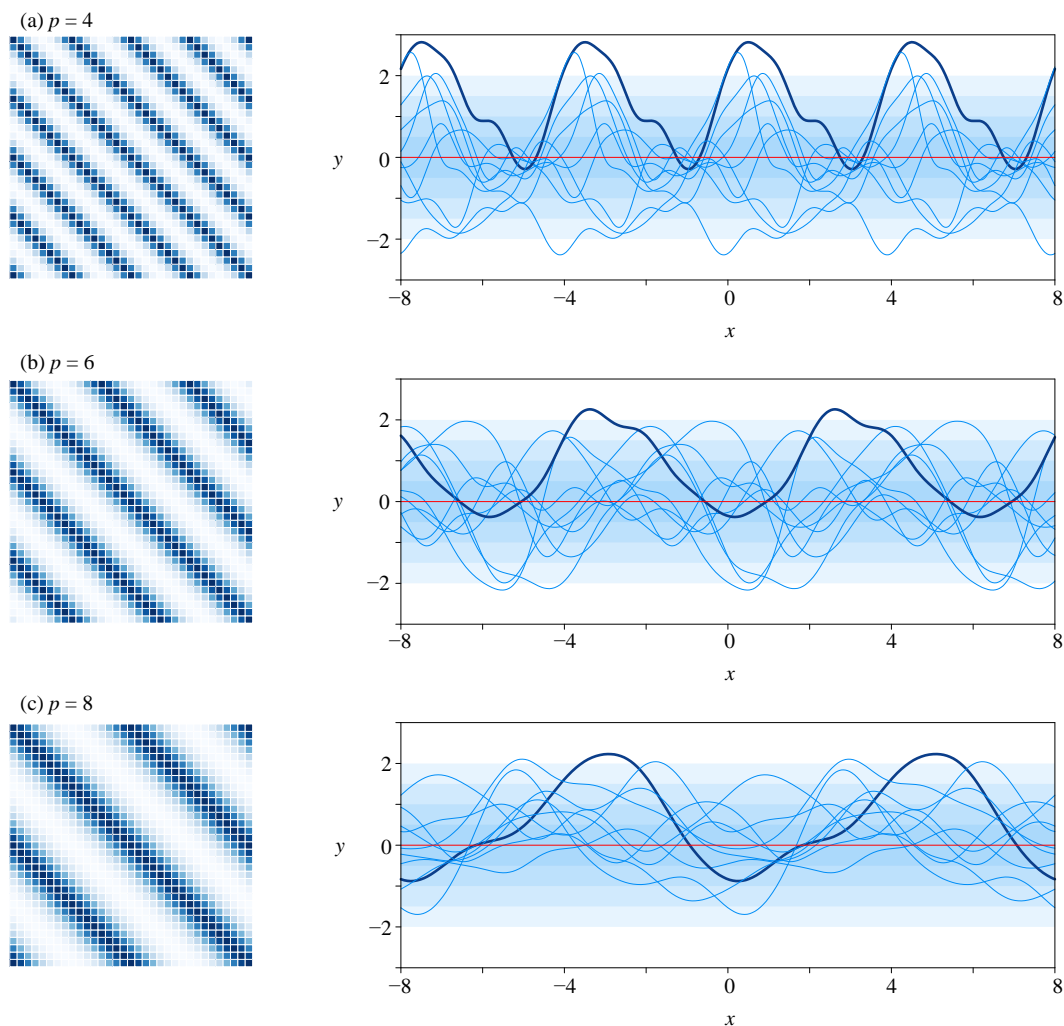
图 29. 无限大的先验协方差矩阵，周期核

下面着重介绍一下参数 p 周期核影响。

如图 30 所示，随着参数 p 增大，采样曲线的波动周期不断变长。

周期核的关键特点在于它引入了正弦函数，使得核函数对周期性变化非常敏感。当输入点在周期上相隔较短的距离时，核函数的值较大，表示这些点在函数空间中具有相似的输出。而当输入点在周期上相隔较远时，核函数的值较小，表示它们在输出上有较大的差异。

周期核常用于建模具有明显周期性结构的时间序列数据或周期性变化的信号。选择合适的周期和长度尺度参数是使用周期核的关键，这样可以使高斯过程模型更好地捕捉数据中的周期性模式。

图 30. 高斯核参数 σ 对先验协方差和采样影响, $l = 1$

核函数的组合

在高斯过程中，我们还可以通过加法或乘法叠加不同核函数以便构建更复杂、更灵活的核函数。这种方式使得高斯过程模型能够更好地适应不同类型的数据模式。

比如，通过乘法获得两个核函数的乘积：

$$\kappa(x_i, x_j) = \text{cov}(y_i, y_j) = \kappa_1(x_i, x_j) \cdot \kappa_2(x_i, x_j) \quad (13)$$

这种方式常用于组合两个核函数的优点，例如结合具有周期性和长度尺度的核函数，以适应同时存在周期性和趋势性的数据。当然，我们也可以将更多不同类型核函数通过乘积方式组合起来。

再比如，通过加法获得两个核函数的和：

$$\kappa(x_i, x_j) = \text{cov}(y_i, y_j) = \kappa_1(x_i, x_j) + \kappa_2(x_i, x_j) \quad (14)$$

通过加法叠加，可以将两个核函数的特性相加，得到新的核函数。这种方式常用于处理数据中不同尺度的变化，例如同时存在高频和低频成分的数据。

下面举三个例子，用乘法组合两个不同的核函数。

图 31 所示为高斯核和线性核的乘积，即高斯核 \times 线性核。图 32 所示为高斯核 \times 线性核先验协方差矩阵及采样曲线。

线性核部分可以处理数据的线性趋势，而高斯核部分则引入了非线性特性，能够捕捉数据中的复杂模式和局部关系。在实际应用中，这种组合核函数常用于处理同时存在线性和非线性结构的数据。例如，当数据在整体上呈线性趋势，但在局部存在一些非线性的波动或变化时，使用线性核和高斯核的乘积可以更好地拟合数据。

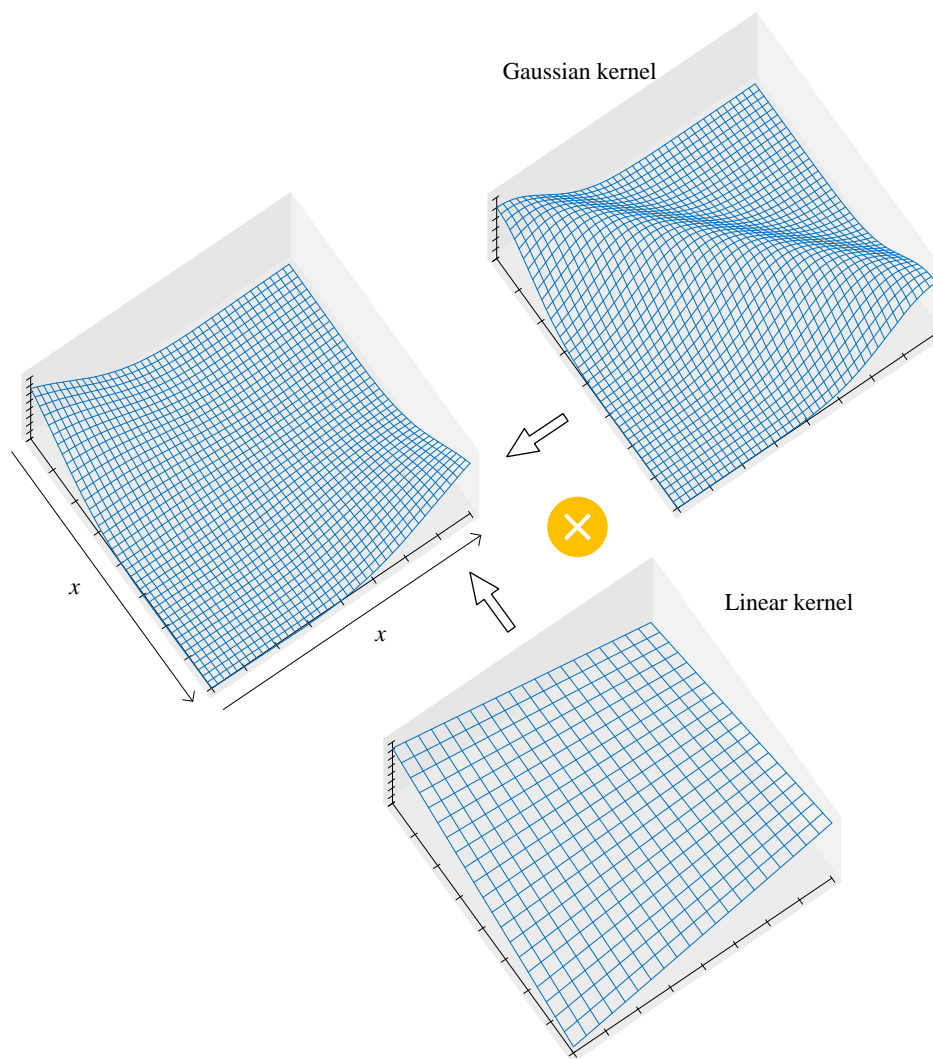


图 31. 高斯核和线性核的乘积

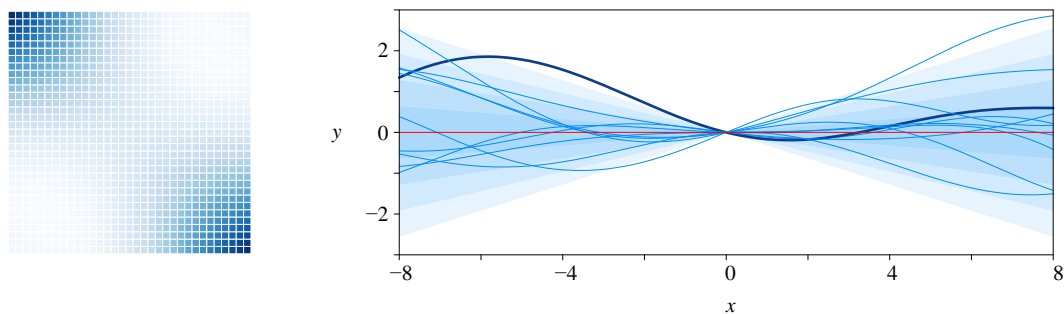


图 32. 高斯核和线性核的乘积，先验协方差矩阵和采样

图 33 所示为高斯核和周期核的乘积，即高斯核 \times 周期核。图 34 所示为高斯核 \times 周期核先验协方差矩阵及采样曲线。这样的组合核函数结合了周期性和非周期性的特性。周期核部分能够捕捉数据中的周期性结构，而高斯核部分引入了非周期性的平滑性，使模型对整体趋势有更好的拟合能力。

在实际应用中，这种组合核函数常用于建模具有明显周期性变化，同时又包含一些噪声或非周期性成分的数据。例如，对于时间序列数据，可能存在明显的季节性变化（周期性），同时受到其他因素的影响（非周期性）。通过将周期核和高斯核进行乘积，模型能够更全面地考虑这两种特性，提高对复杂数据模式的拟合能力。

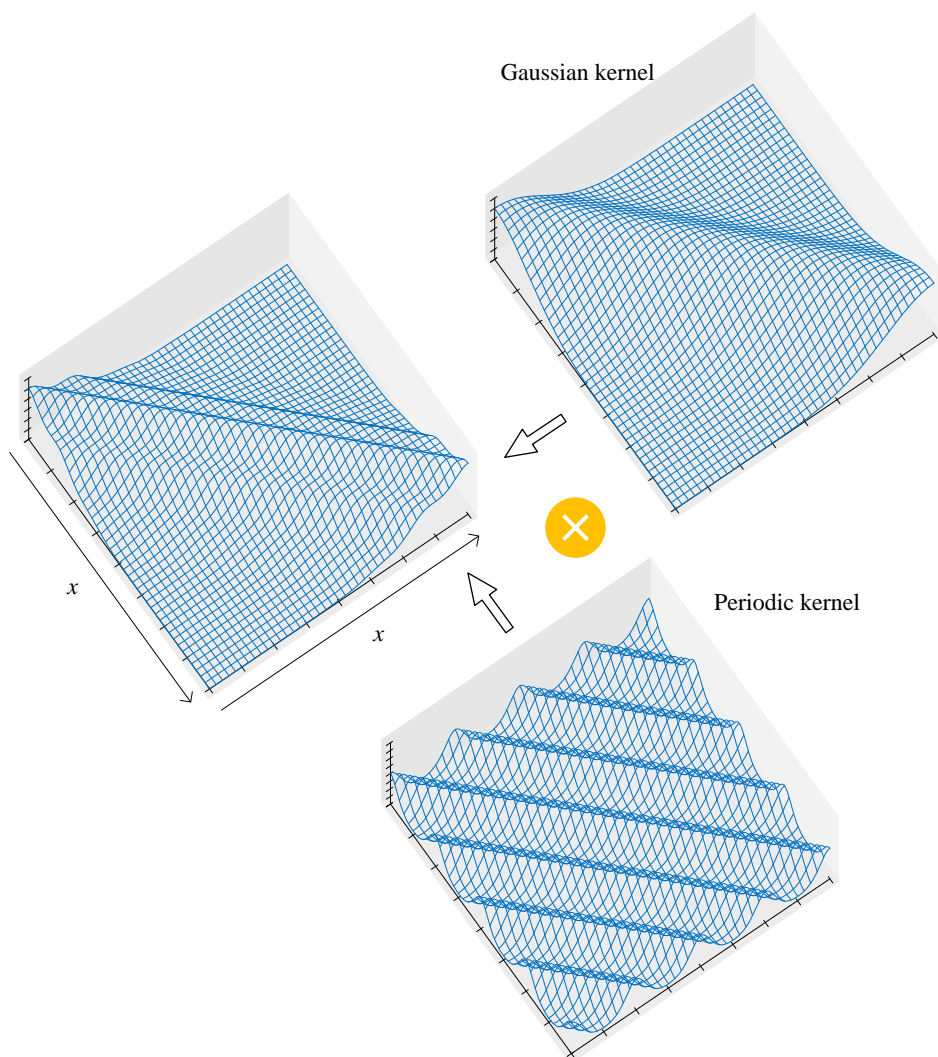


图 33. 高斯核和周期核的乘积

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger: <https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

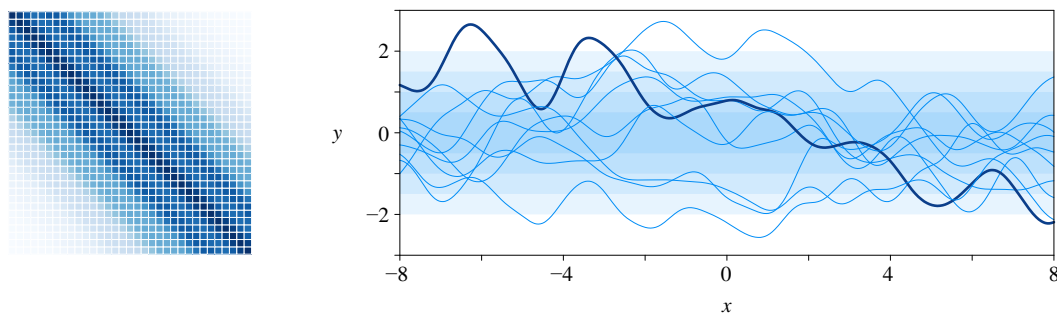


图 34. 高斯核和周期核的乘积，先验协方差矩阵和采样

图 35 所示为线性核和周期核的乘积，即线性核 \times 周期核。图 36 所示为线性核 \times 周期核先验协方差矩阵及采样曲线。这样的组合核函数同时包含了周期性和线性趋势的特性。周期核部分捕捉了数据中的周期性结构，而线性核部分用于处理数据的线性趋势。通过这种组合，模型可以更灵活地适应同时存在周期性和线性结构的数据。

在实际应用中，这种组合核函数常用于处理同时具有周期性和整体线性趋势的数据，例如时间序列数据中同时存在季节性变化和总体趋势的情况。

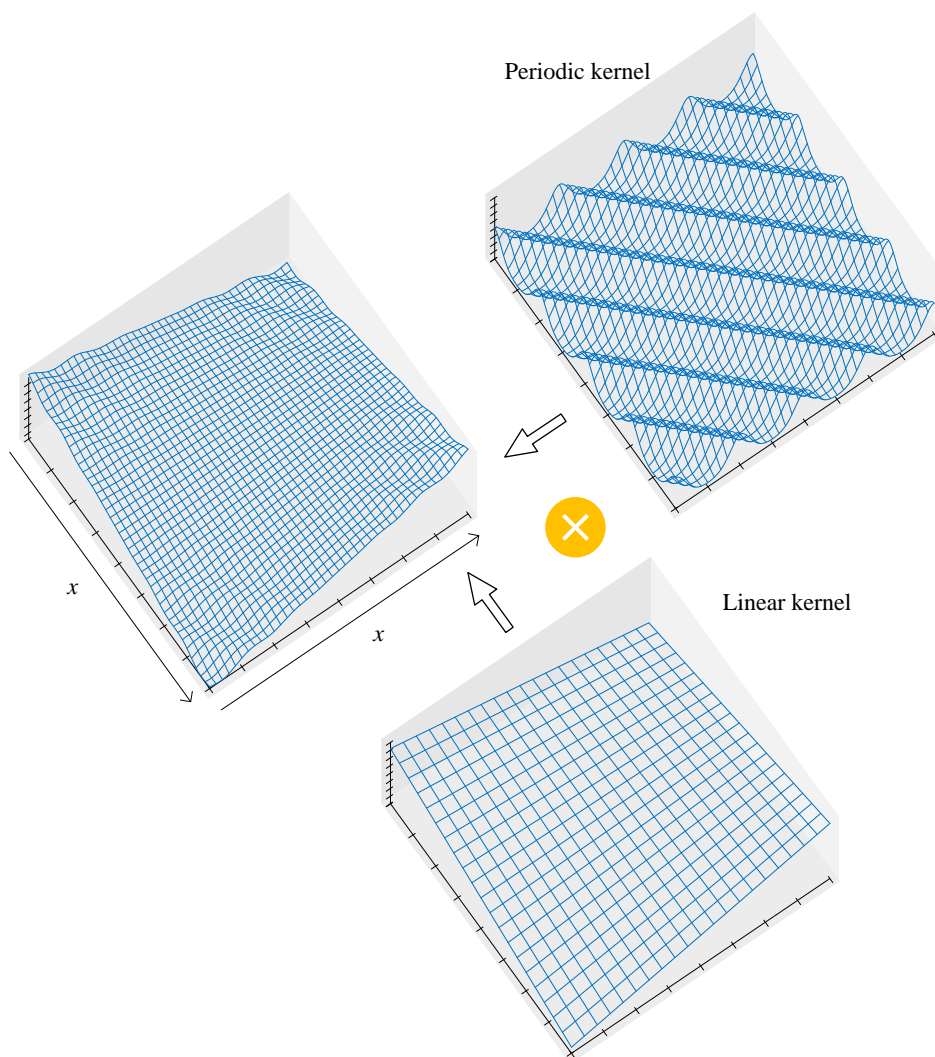


图 35. 线性核和周期核的乘积

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

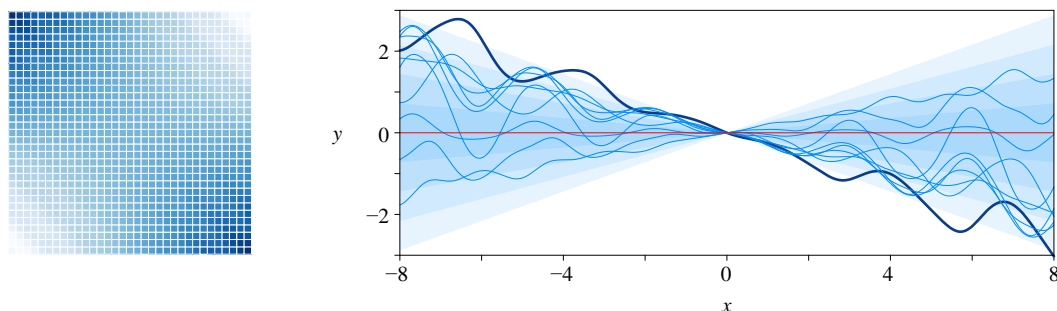


图 36. 线性核和周期核的乘积，先验协方差矩阵和采样

《机器学习》在讲解支持向量机核技巧时还会提到不同的核函数。此外，《机器学习》还会介绍如何用 Scikit-learn 库中高斯过程工具完成回归和分类。

大家想要深入学习高斯过程，请参考如下开源图书 *Gaussian Processes for Machine Learning*:

<https://gaussianprocess.org/gpml/>

这篇博士论文中专门介绍了不同核函数的叠加：

<https://www.cs.toronto.edu/~duvenaud/thesis.pdf>

作者认为下面这篇文章解释高斯过程做的交互设计最佳，这篇文章给了作者很多可视化方面的启发：

<https://distill.pub/2019/visual-exploration-gaussian-processes/>

高斯过程可谓高斯分布和贝叶斯定理的完美结合体。本章的关键是理解高斯过程算法原理。希望大家学完这章后，能够掌握如何用高斯核函数构造协方差矩阵，并计算后验分布。本章最后介绍高斯过程中可能用到的更多核函数。

此外，《机器学习》一册还会再介绍高斯过程，我们会用 Scikit-learn 中高斯过程工具完成回归和分类两种不同问题。