

10

Bivariate Gaussian Distribution

二元高斯分布

椭圆的极致应用



自然之书是用数学语言写成的，符号是三角形、圆形和其他几何图形；不理解几何图形，别想读懂自然之书；没有它们，我们只能在黑暗的迷宫中徘徊不前。

The book of nature is written in mathematical language, and the symbols are triangles, circles and other geometrical figures, without whose help it is impossible to comprehend a single word of it; without which one wanders in vain through a dark labyrinth.

—— 伽利略·伽利莱 (Galilei Galileo) | 意大利物理学家、数学家及哲学家 | 1564 ~ 1642



```
◀ scipy.stats.multivariate_normal() 多元高斯分布
◀ scipy.stats.multivariate_normal.pdf() 多元高斯分布 PDF 函数
◀ scipy.stats.multivariate_normal.cdf() 多元高斯分布 CDF 函数
◀ ax.plot_wireframe() 绘制线框图
◀ matplotlib.pyplot.axhline() 绘制水平线
◀ matplotlib.pyplot.axvline() 绘制竖直线
◀ matplotlib.pyplot.contour() 绘制等高线图
◀ matplotlib.pyplot.contourf() 绘制填充等高线图
◀ matplotlib.patches.Rectangle() 绘制长方形
```



10.1 二元高斯分布：看见椭圆

概率密度函数

二元高斯分布 (bivariate Gaussian distribution)，也叫二元正态分布 (bivariate normal distribution)，它的概率密度函数 $f_{X,Y}(x,y)$ 解析式如下：

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \times \exp \left(\underbrace{\frac{-1}{2(1-\rho_{X,Y}^2)} \left(\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho_{X,Y} \left(\frac{x-\mu_X}{\sigma_X} \right) \left(\frac{y-\mu_Y}{\sigma_Y} \right) + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right)}_{\text{Ellipse}} \right) \quad (1)$$

其中， μ_X 和 μ_Y 分别为随机变量 X 、 Y 的期望值， σ_X 和 σ_Y 分别为随机变量 X 、 Y 的均方差， $\rho_{X,Y}$ 为 X 和 Y 线性相关系数。观察 (1)，显然 $\rho_{X,Y}$ 不能取值 ± 1 ；否则，分母为 0 取值区间为 $(-1, 1)$ 。

此外，丛书之前反复提到二元高斯分布和椭圆的关系。我们在 (1) 中已经看到了椭圆解析式。这种椭圆解析式形式正是我们在《数学要素》第 9 章讲过的特殊类型。

PDF 曲面形状

给定如下条件：

$$\mu_X = 0, \mu_Y = 0, \sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75 \quad (2)$$

绘制满足条件的二元正态分布密度函数曲面，具体如图 1 所示。容易发现， μ_X 和 μ_Y 决定曲面中心所在位置； σ_X 和 σ_Y 影响曲面在 x 和 y 方向的形状。而 $\rho_{X,Y}$ 似乎提供了曲面的扭曲。

下面，我们从几个侧面来深入二元高斯分布 PDF $f_{X,Y}(x,y)$ 曲面。

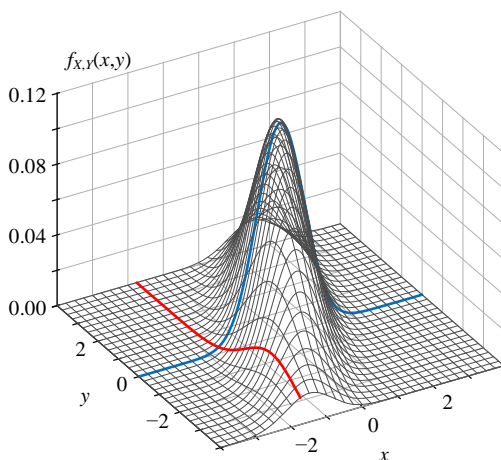


图 1. 二元高斯分布 PDF 函数曲面 $f_{X,Y}(x,y)$ ， $\sigma_X = 1$ ， $\sigma_Y = 2$ ， $\rho_{X,Y} = 0.75$

沿 x 剖面线

图 2 所示为 $f_{X,Y}(x,y)$ 曲面沿 x 方向的剖面线，以及这些曲线在 xz 平面上的投影。这些曲线，相当于是 (1) 中 y 取定值。比如 $y = 0$ 时，曲线的解析式：

$$f_{X,Y}(x, y=0) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \times \exp\left(-\frac{1}{2(1-\rho_{X,Y}^2)}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \frac{2\rho_{X,Y}\mu_Y}{\sigma_Y}\left(\frac{x-\mu_X}{\sigma_X}\right) + \left(\frac{\mu_Y}{\sigma_Y}\right)^2\right)\right) \quad (3)$$

这条曲线，我们看到的是一元正态分布的影子。注意，这个曲线和横轴围成的图形面积并不为 1，这个图形的面积是边缘 PDF $f_Y(y=0)$ 。

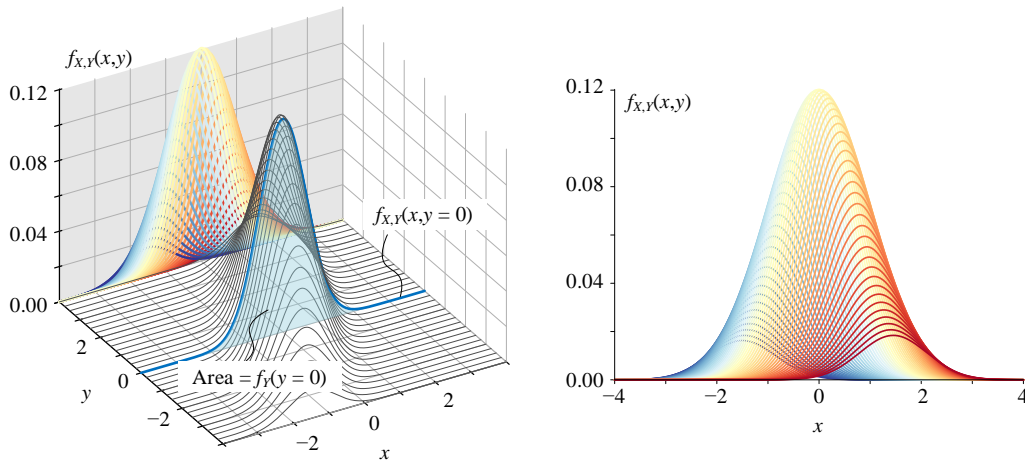


图 2. PDF 函数曲面 $f_{X,Y}(x,y)$ ，沿 x 方向的剖面线， $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$

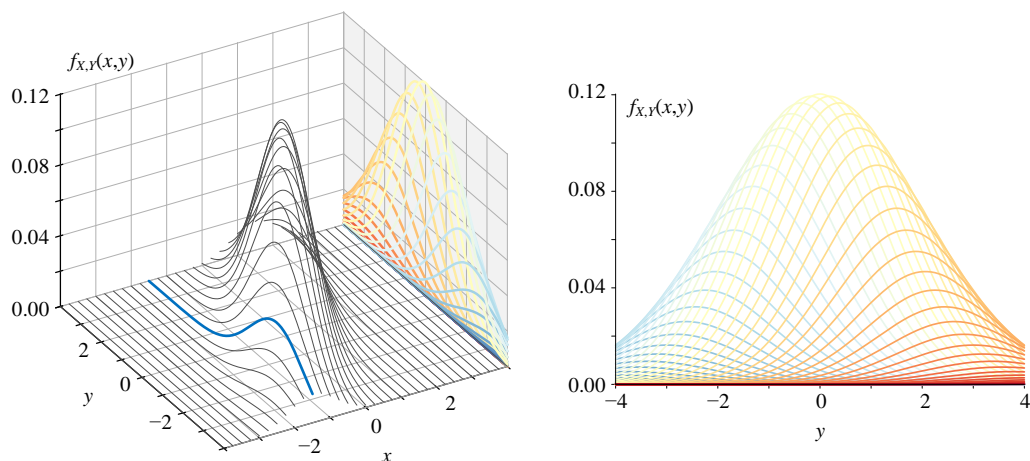
有读者可能会想，如果我们可以得到 $y = 0$ 时边缘 PDF $f_Y(y=0)$ 的具体值，就可以通过下式得到条件概率 $f_{X|Y}(x|y=0)$ 函数：

$$f_{X|Y}(x|y=0) = \frac{f_{X,Y}(x, y=0)}{f_Y(y=0)} \quad (4)$$

我们实际上可以获得上式的解析式，这是本书第 13 章要讲解的内容。

沿 y 剖面线

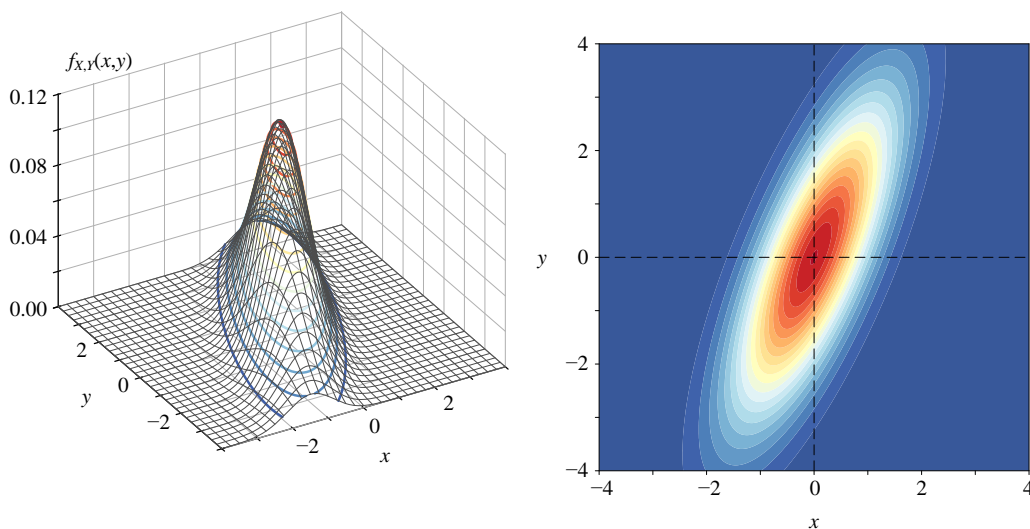
图 3 所示为 $f_{X,Y}(x,y)$ 曲面沿 y 方向的剖面线，以及这些曲线在 yz 平面上的投影。曲线相当于 y 取定值，联合 PDF 随 y 变化。

图 3. PDF 函数曲面 $f_{X,Y}(x,y)$, 沿 y 方向的剖面线, $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$

等高线

图 4 所示为 $f_{X,Y}(x,y)$ 曲面三维等高线。很明显，我们已经从等高线中明显地看到椭圆。图 4 所示平面等高线图印证了有关椭圆的结论。大家从图 4 (b) 中看到一系列同心旋转椭圆。这并不奇怪，因为 (1) 中 $\exp()$ 函数中蕴含着一个椭圆解析式。

这也就是为什么高斯分布被称作是椭圆分布 (elliptical distribution) 的一种。本章后续将会揭开高斯分布和椭圆的更多联系。

图 4. PDF 函数曲面 $f_{X,Y}(x,y)$, 空间等高线和平面填充等高线, $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$

相关性系数

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

为了方便大家了解相关性系数对二元高斯分布 PDF 的影响，设定 $\sigma_X = 1, \sigma_Y = 1$ ，如图 5 所示，不同相关性系数，二元高斯分布 PDF 曲面和等高线。

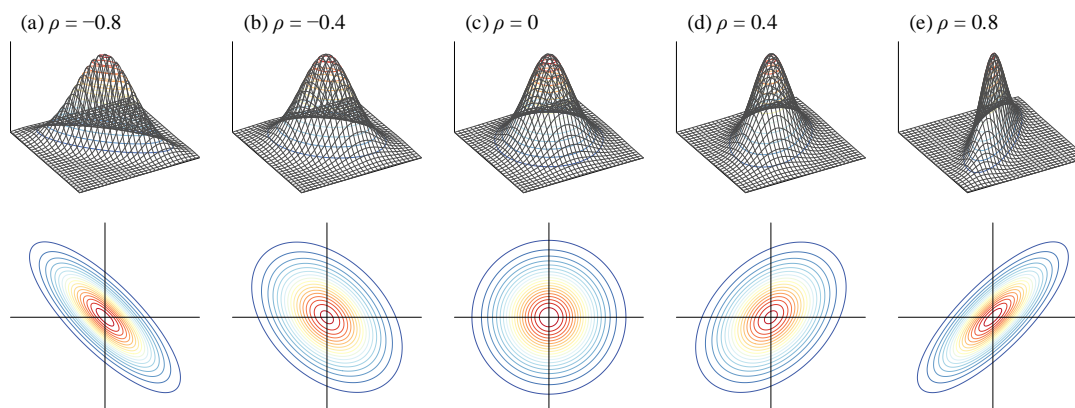


图 5. 不同相关性系数，二元高斯分布 PDF 曲面和等高线， $\sigma_X = 1, \sigma_Y = 1$

质心

如图 6 所示，质心仅仅影响曲面中心位置。

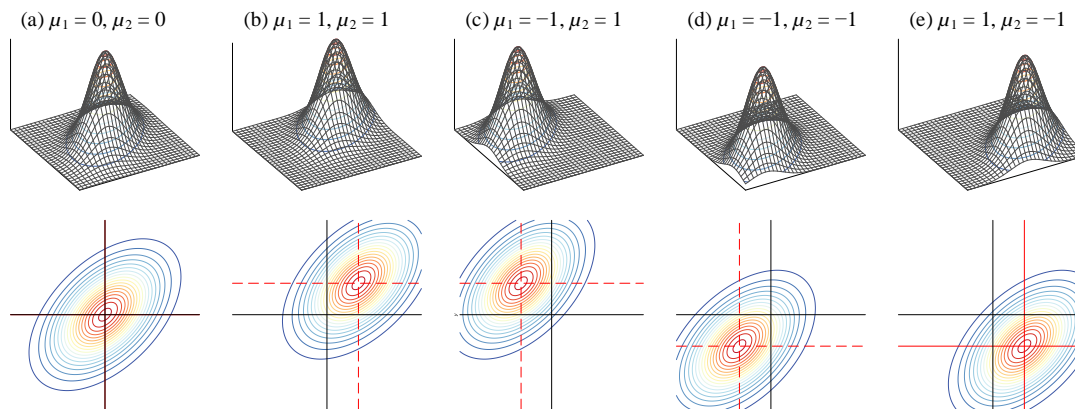


图 6. 不同质心位置，二元高斯分布 PDF 曲面和等高线， $\sigma_X = 1, \sigma_Y = 1$



Bk5_Ch09_01.py 绘制本节图像。

10.2 边缘分布：一元高斯分布

边缘分布

大家可能也已经注意到，不考虑 Y 的时候， μ_X 和 σ_X 仅仅是描述随机变量 X 的参数。也就是说，有了和这两个参数，我们就可以写出 X 的边缘 PDF $f_X(x)$ —— 一元高斯分布概率密度函数：

$$f_X(x) = \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x - \mu_X}{\sigma_X}\right)^2\right) \quad (5)$$

同理， μ_Y 和 σ_Y 是描述随机变量 Y 的参数，对应写出 Y 的边缘 PDF $f_Y(y)$ ：

$$f_Y(y) = \frac{1}{\sigma_Y \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{y - \mu_Y}{\sigma_Y}\right)^2\right) \quad (6)$$

在图 4 平面等高线基础上添加 $f_X(x)$ 和 $f_Y(y)$ 边缘 PDF 图像子图，我们便得到图 7。

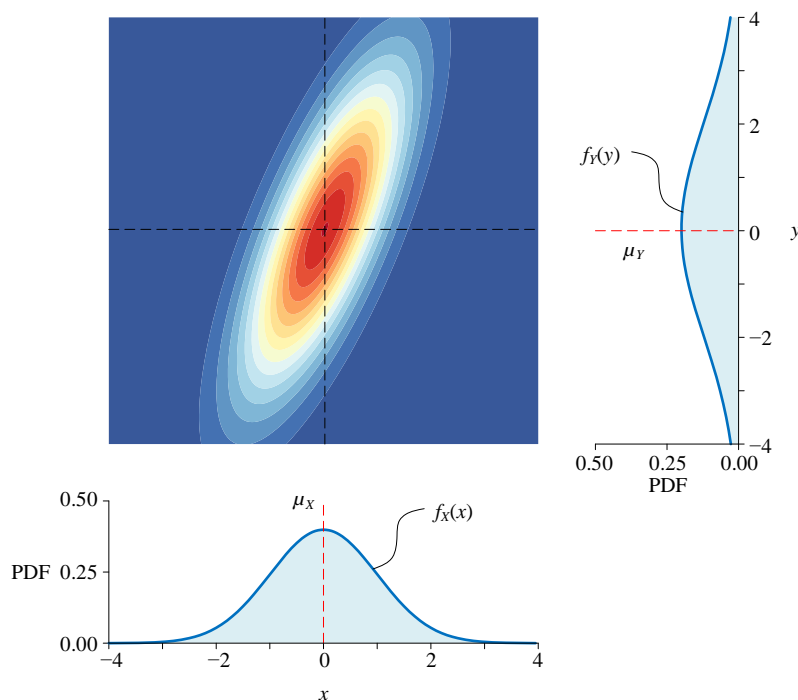


图 7. 二元高斯分布 PDF 和边缘 PDF, $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$

偏积分求边缘分布 PDF

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

下面，以 X 的边缘分布概率密度函数 $f_X(x)$ 为例证明二元高斯分布 PDF“偏积分”得到一元高斯分布 PDF。

连续随机变量 X 的边缘分布概率密度函数 $f_Y(y)$ 可以通过 $f_{X,Y}(x,y)$ 对 x 偏积分得到，即：

$$f_Y(y) = \overbrace{\int_{-\infty}^{+\infty} f_{X,Y}(x,y) dx}^{\text{Eliminate } x} \quad (7)$$

令，

$$Q(x,y) = \frac{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho_{X,Y}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}{(1-\rho_{X,Y}^2)} \quad (8)$$

这样，二元高斯分布可以写成：

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \times \exp\left(\frac{-1}{2}Q(x,y)\right) \quad (9)$$

将 (8) 中 $Q(x,y)$ 写成：

$$\begin{aligned} Q(x,y) &= \frac{\left(\frac{x-\mu_X}{\sigma_X} - \rho_{X,Y}\frac{y-\mu_Y}{\sigma_Y}\right)^2}{(1-\rho_{X,Y}^2)} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 \\ &= \frac{\left(x - \left(\mu_X + \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right)\right)^2}{(1-\rho_{X,Y}^2)\sigma_X^2} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 \end{aligned} \quad (10)$$

令

$$t = t(y) = \mu_X + \rho_{X,Y}\frac{\sigma_X}{\sigma_Y}(y-\mu_Y) \quad (11)$$

可以发现 t 仅仅是 y 的函数，与 x 无关，这样便于积分。

$Q(x,y)$ 进一步整理为：

$$Q(x,y) = \frac{(x-t)^2}{(1-\rho_{X,Y}^2)\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} \quad (12)$$

将 (12) 代入 (9) 得到：

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \times \exp\left(\frac{-1}{2}\left(\frac{(x-t)^2}{(1-\rho_{X,Y}^2)\sigma_X^2}\right)\right) \times \exp\left(\frac{-1}{2}\left(\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right) \quad (13)$$

将 (13) 代入 (7) 得到：

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{+\infty} \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \times \exp\left(\frac{-1}{2}\left(\frac{(x-t)^2}{(1-\rho_{X,Y}^2)\sigma_X^2}\right)\right) \times \exp\left(\frac{-1}{2}\left(\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right) dx \\
&= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \cdot \exp\left(\frac{-1}{2}\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right) \cdot \int_{-\infty}^{+\infty} \exp\left(\frac{-1}{2}\left(\frac{(x-t)^2}{(\sqrt{(1-\rho_{X,Y}^2})\sigma_X)^2}\right)\right) dx
\end{aligned} \tag{14}$$

回忆，我们在微积分部分讲解的如下积分：

$$\int_{-\infty}^{+\infty} \exp\left(\frac{-1}{2}\left(\frac{(x-t)^2}{(\sqrt{(1-\rho_{X,Y}^2})\sigma_X)^2}\right)\right) dx = \sqrt{2\pi}\sqrt{1-\rho_{X,Y}^2}\sigma_X \tag{15}$$

将 (15) 代入 (14)，得到：

$$\begin{aligned}
f_Y(y) &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \cdot \exp\left(\frac{-1}{2}\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right) \sqrt{2\pi}\sqrt{(1-\rho_{X,Y}^2)}\sigma_X \\
&= \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left(\frac{-1}{2}\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)
\end{aligned} \tag{16}$$

再次强调，联合 PDF $f_{X,Y}(x,y)$ 二重积分得到的是概率，也就是曲面体积代表概率。而 $f_{X,Y}(x,y)$ 偏积分得到的还是概率密度，即边缘概率密度 $f_X(x)$ 或 $f_Y(y)$ 。边缘 PDF $f_X(x)$ 和 $f_Y(y)$ 进一步积分才得到概率。

独立

图 8 所示为二元高斯分布参数对 PDF 等高线影响。特别地，当相关性系数 $\rho_{X,Y}$ 为 0 时：

$$\begin{aligned}
f_{X,Y}(x,y) &= \frac{1}{2\pi\sigma_1\sigma_2} \times \exp\left(\frac{-1}{2}\left(\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right)\right) \\
&= \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(\frac{-1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right) \times \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left(\frac{-1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right) \\
&= f_X(x)f_Y(y)
\end{aligned} \tag{17}$$

观察图 7 (b)、(e)、(h)，我们发现椭圆等高线为正椭圆。

注意，相关系数为 0 是两变量独立的必要非充分条件。相关系数反映的是两变量间的线性关系，但是变量间除了线性关系还有其它关系，这时候线性相关系数就不能作为度量尺度。本书第 15 章将专门讲解相关性。

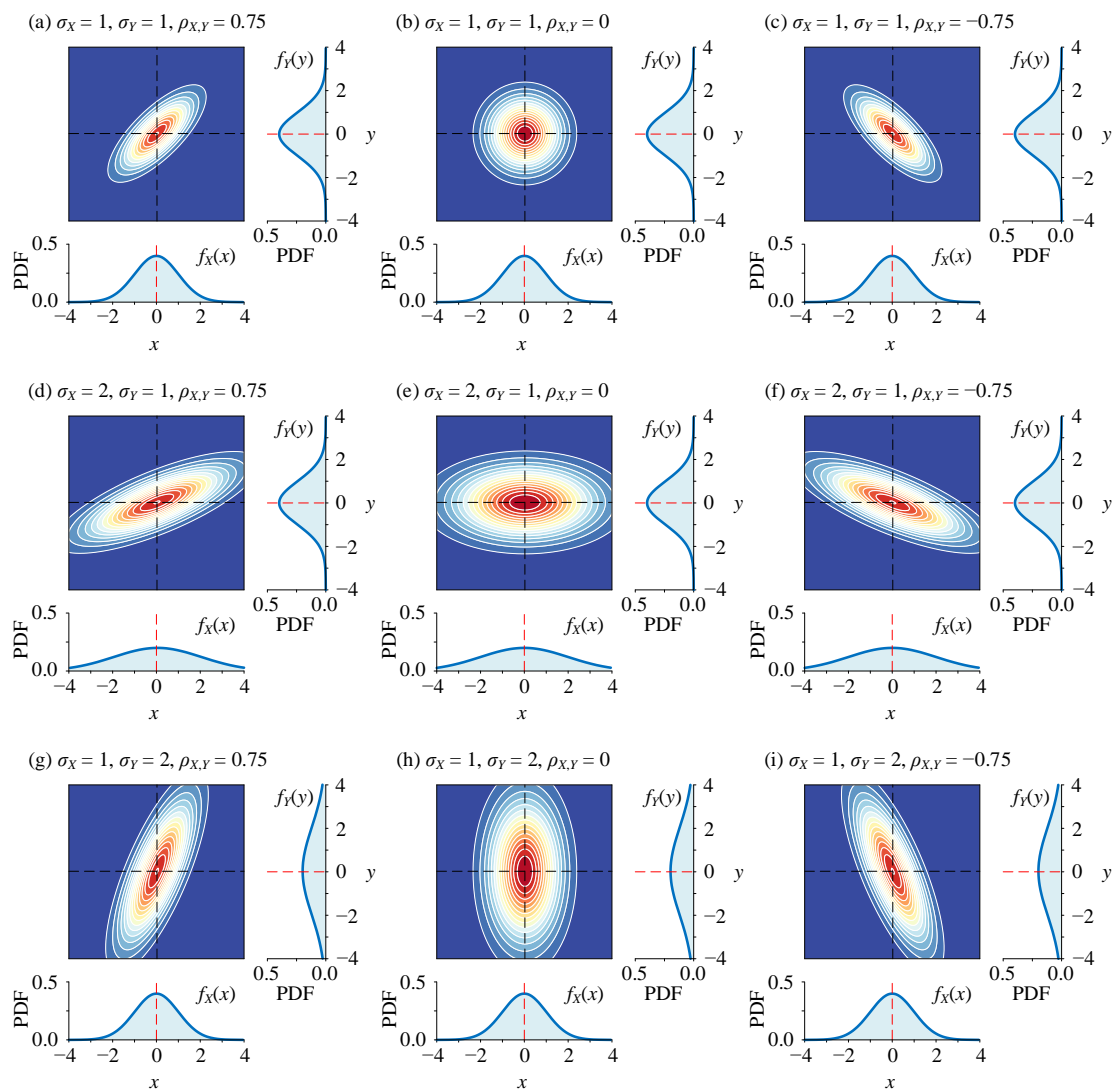


图 8. 二元高斯分布参数对 PDF 等高线影响



Bk5_Ch09_02.py 绘制本节图像。请大家自行调整分布参数。

10.3 累积概率密度：概率值

二元高斯分布的累积分布函数 CDF $F_{X,Y}(x,y)$ 是对 PDF $f_{X,Y}(x,y)$ 的二重积分：

$$F_{X,Y}(x,y) = \int_{-\infty}^y \int_{-\infty}^x f_{X,Y}(s,t) ds dt \quad (18)$$

图 9 所示为二元高斯分布累积分布函数 CDF 曲面。

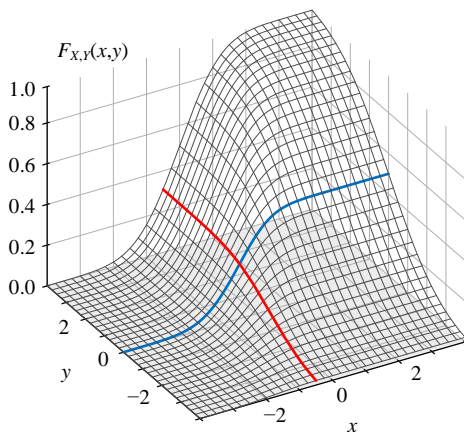
本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

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>

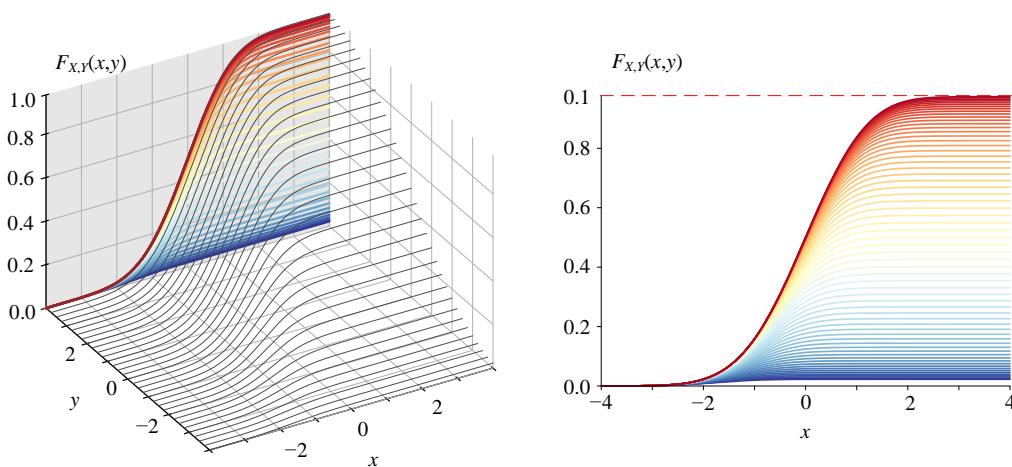
本书配套微课视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

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

图 9. 二元高斯分布累积函数 CDF 曲面, $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$

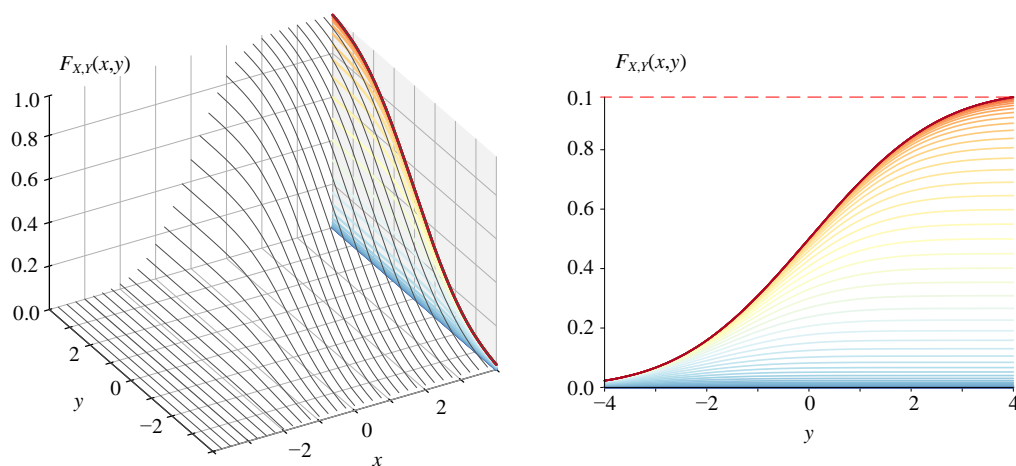
沿 x 剖面线

和上一节一样，从几个侧面来深入二元高斯分布 CDF 曲面 $F_{X,Y}(x,y)$ 。图 10 所示为 $F_{X,Y}(x,y)$ 曲面沿 x 方向的剖面线，以及这些曲线在 xz 平面上的投影。

图 10. CDF 曲面 $F_{X,Y}(x,y)$, 沿 x 方向的剖面线, $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$

沿 y 剖面线

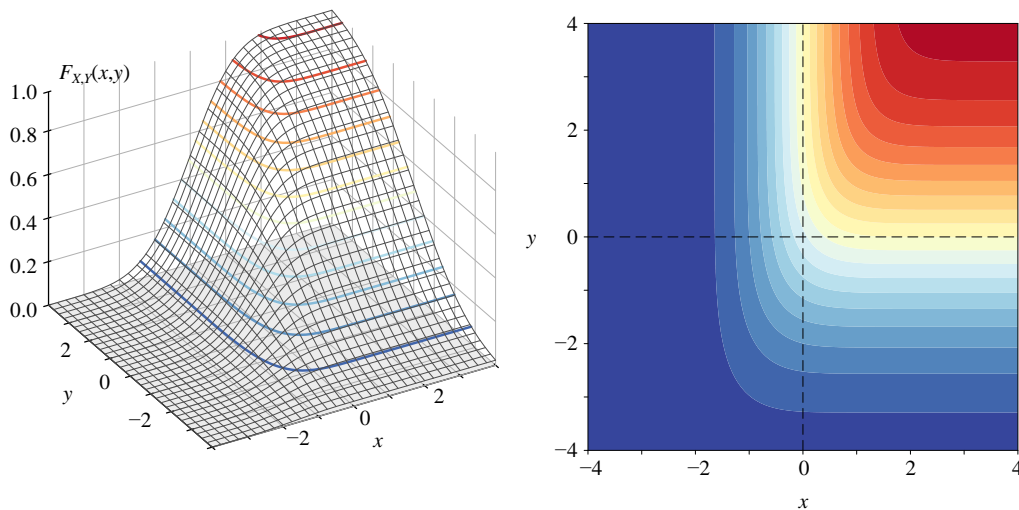
图 11 所示为 $F_{X,Y}(x,y)$ 曲面沿 y 方向的剖面线，以及这些曲线在 yz 平面上的投影。

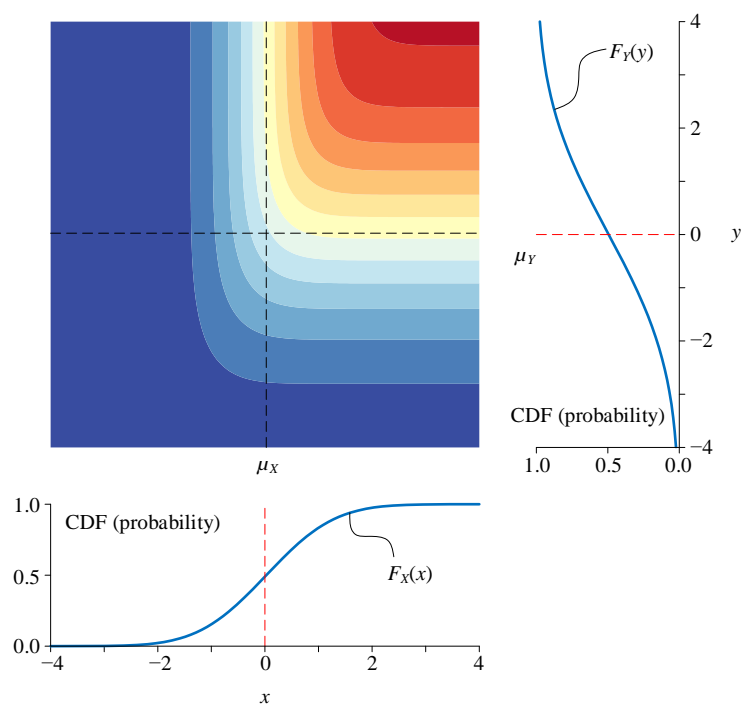
图 11. CDF 曲面 $F_{X,Y}(x,y)$, 沿 y 方向的剖面线, $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$

等高线

图 12 所示为 CDF 函数曲面 $F_{X,Y}(x,y)$, 空间等高线和平面填充等高线。

请大家修改上一节代码绘制本节图像。大家只需要把 `scipy.stats.multivariate_normal.pdf()` 换成 `scipy.stats.multivariate_normal.cdf()` 函数。

图 12. CDF 函数曲面 $F_{X,Y}(x,y)$, 空间等高线和平面填充等高线, $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$

图 13. CDF 函数曲面 $F_{x,y}(x,y)$ 平面填充等高线，边缘概率分布

10.4 用椭圆解剖二元高斯分布

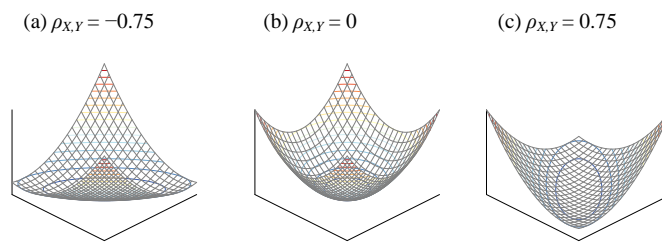
二次曲面

这一节我们对二元高斯分布和椭圆的关系进行定量研究。大家已经在 (1) 看到了椭圆的影子。利用 (8) 中定义的 $G(x,y)$ 。将 (8) 代入 (1)，得到：

$$f_{x,y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{x,y}^2}} \times \exp\left(\frac{-1}{2}G(x,y)\right) \quad (19)$$

图 14 所示为 $G(x,y)$ 代表的几种曲面。

但是，对于二元高斯分布来说，如果 PDF 解析式存在，相关性的取值范围为 $(-1, 1)$ ，此时协方差矩阵为正定。请大家思考如果，协方差矩阵为半正定， $G(x,y)$ 曲面的形状是什么？

图 14. $G(x, y)$ 代表的几种曲面

椭圆

令 $G(x, y) = 1$ ，当 $\rho_{X,Y}$ 在 $(-1, 1)$ 变化时，我们便得到椭圆的解析式：

$$\frac{1}{(1-\rho_{X,Y}^2)} \left(\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho_{X,Y} \left(\frac{x-\mu_X}{\sigma_X} \right) \left(\frac{y-\mu_Y}{\sigma_Y} \right) + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right) = 1 \quad (20)$$

(μ_1, μ_2) 确定椭圆中心位置， σ_1 、 σ_2 和 ρ 三者共同决定椭圆长短轴长度和旋转角度。

丛书《数学要素》第 9 章介绍过，形如 (20) 解析式的椭圆有重要的特点——椭圆和长 $2\sigma_X$ 、宽 $2\sigma_Y$ 的矩形相切。

图 15 给出的矩形框中心位于 (μ_X, μ_Y) ，矩形框长度为 $2\sigma_X = 2$ ，矩形框宽度为 $2\sigma_Y = 4$ 。相关性系数 $\rho_{X,Y}$ 的变化范围为 $[-0.9, 0.9]$ 。

相关性系数 ρ 大于 0 时，即正相关，椭圆主轴朝东北方向旋转；相关性系数 ρ 小于 0 时，即负相关，椭圆主轴朝西北方向旋转；特别提醒读者注意的是，当相关性系数 ρ 为 0 时，椭圆为正椭圆。

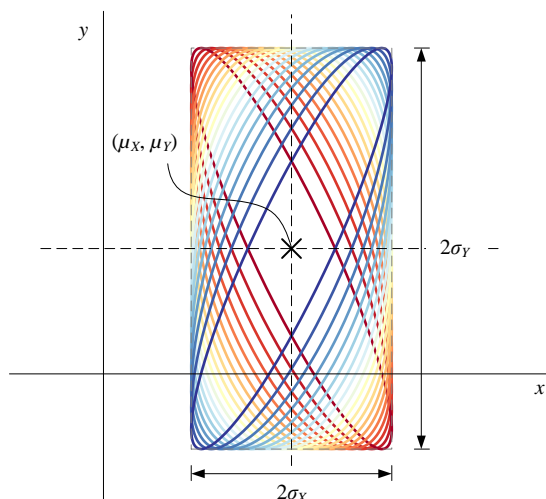
图 15. 椭圆和中心在 (μ_X, μ_Y) 长 $2\sigma_X$ 、宽 $2\sigma_Y$ 的矩形相切

图 16 所示为三种均方差 σ_X 、 σ_Y 大小不同的情况，和矩形相切的椭圆随着相关性系数变化情况。

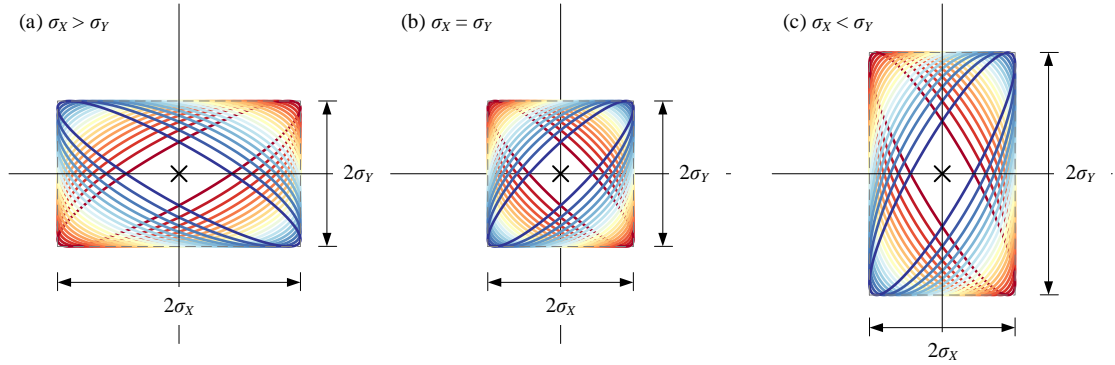


图 16. 三种均方差 σ_X 、 σ_Y 大小不同的情况

四个切点

椭圆和矩形有四个切点，下面我们来求解这四个切点的位置。考虑特殊情况 $\mu_X = 0, \mu_Y = 0$ ，(20) 可以简化为：

$$\frac{1}{(1-\rho_{X,Y}^2)} \left(\left(\frac{x}{\sigma_X} \right)^2 - \frac{2\rho_{X,Y}}{\sigma_X\sigma_Y} xy + \left(\frac{y}{\sigma_Y} \right)^2 \right) = 1 \quad (21)$$

将 $y = \sigma_Y$ 代入 (21)，得到：

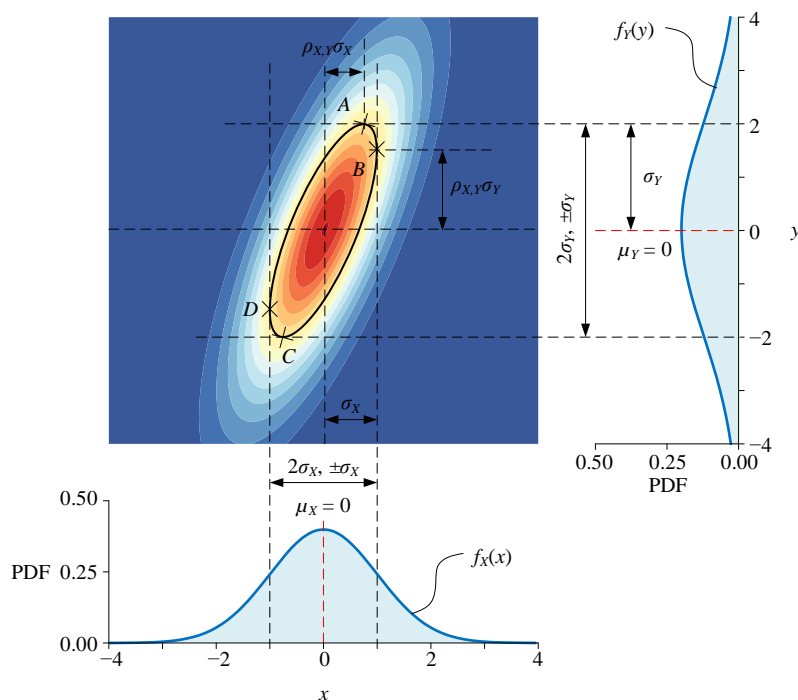
$$\left(\frac{x}{\sigma_X} - \rho_{X,Y} \right)^2 = 0 \quad (22)$$

这样我们便得到一个切点：

$$\begin{cases} x = \rho_{X,Y}\sigma_X \\ y = \sigma_Y \end{cases} \quad (23)$$

同理，获得所有四个切点 A 、 B 、 C 、 D 的具体位置：

$$A(\rho_{X,Y}\sigma_X, \sigma_Y), B(\sigma_X, \rho_{X,Y}\sigma_Y), C(-\rho_{X,Y}\sigma_X, -\sigma_Y), D(-\sigma_X, -\rho_{X,Y}\sigma_Y) \quad (24)$$

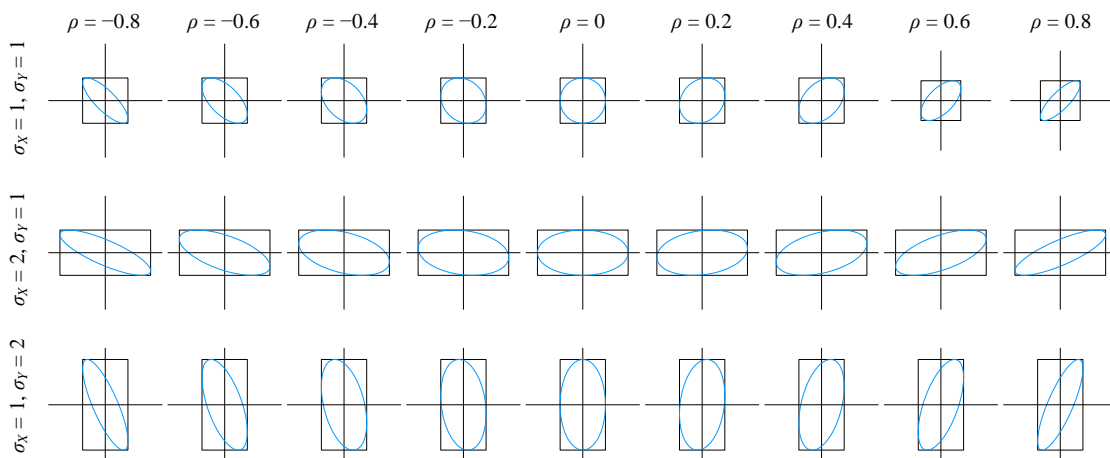
图 17. 二元高斯分布 PDF 和边缘 PDF, $\sigma_X = 1, \sigma_Y = 2, \rho_{X,Y} = 0.75$

椭圆和矩形

μ_X 和 μ_Y 均不为 0 的一般情况，四个切点的位置平移 (μ_X, μ_Y) ，为：

$$\begin{aligned} A(\mu_X + \rho_{X,Y}\sigma_X, \mu_Y + \sigma_Y), \quad B(\mu_X + \sigma_X, \mu_Y + \rho_{X,Y}\sigma_Y), \\ C(\mu_X - \rho_{X,Y}\sigma_X, \mu_Y - \sigma_Y), \quad D(\mu_X - \sigma_X, \mu_Y - \rho_{X,Y}\sigma_Y) \end{aligned} \quad (25)$$

图 18 所示为椭圆和矩形切点位置随 σ_X 、 σ_Y 、 $\rho_{X,Y}$ 变化关系，请大家自行总结规律。

图 18. 椭圆和矩形切点随 σ_X 、 σ_Y 、 $\rho_{X,Y}$ 变化关系



Bk5_Ch09_03.py 绘制图 18。

椭圆形状

再怎么强调椭圆和高斯分布的关系也不为过。图 19 这个旋转椭圆的位置、形状、旋转角度等信息，蕴含着高斯分布的中心 (μ_X, μ_Y) 、均方差 σ_X 和 σ_Y 、相关性系数 $\rho_{X,Y}$ 。也就是说，一个二元高斯分布可以用一个椭圆来代表。

更重要的是，还有很多有关的椭圆的性质值得我们挖掘。设定椭圆的中心为 C ，图 19 所示两个椭圆，蓝色椭圆上所有点到代入 (8) 等于 1，类似一元高斯分布中的 $\pm\sigma$ ；而更大一点的红色椭圆所有点代入 (8) 等于 4，平方根为 2，类似一元高斯分布中的 $\pm 2\sigma$ 。平方根正是《矩阵力量》第 20 章讲过马氏距离。实际上，(20) 中 $G(x, y) = 1$ 意味着马氏距离为 1。本章后文将稍微回顾马氏距离，本书第 23 章还要深入讲解马氏距离。

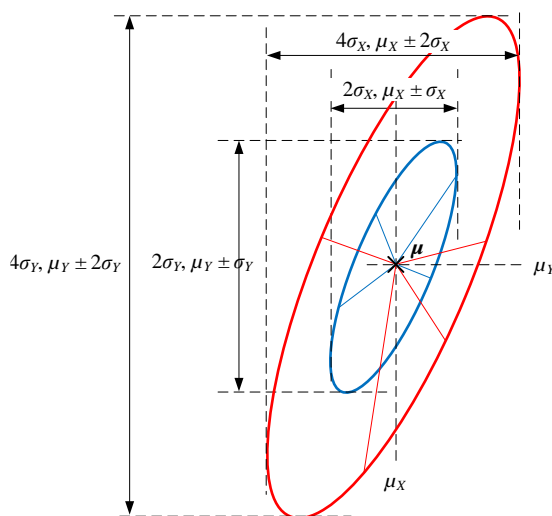


图 19. 两个椭圆

图 20 中的浅蓝色直角三角形的两条直角边长度分别是 $\rho_{X,Y}\sigma_Y$ 、 σ_X ，其中 θ 角的正切值为：

$$\tan \theta = \frac{\rho_{X,Y}\sigma_Y}{\sigma_X} \quad (26)$$

图 20 所示 AC 线段、 BD 线段和条件概率、线性回归有着直接联系。本书第 13 章将专门讲解高斯分布条件概率。

图 20 中两条红色线为椭圆的长轴和短轴，这两条直线又和主成分分析有着密切的关系。这是本书第 25 章要探讨的内容。

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

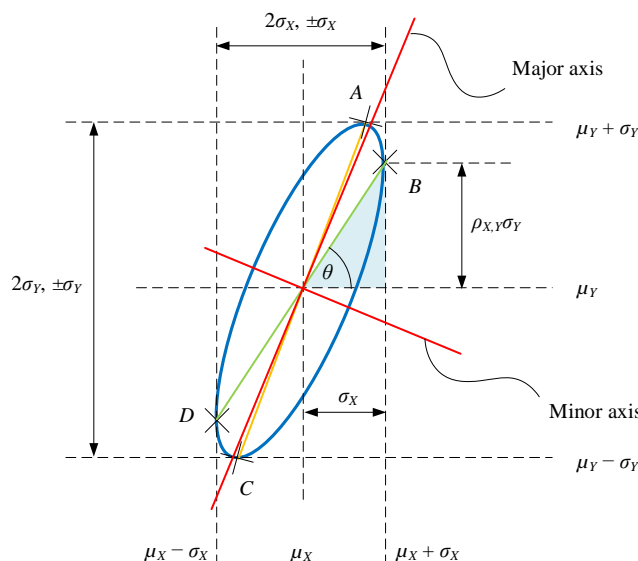


图 20. 椭圆中的四条直线

10.5 以鸢尾花数据为例：不考虑分类标签

本节和下一节用二元高斯分布估计鸢尾花花萼长度 X_1 、花萼宽度 X_2 的联合概率密度函数 $f_{X_1, X_2}(x_1, x_2)$ 。相信大家还记得我们在本书第 7 章采用 KDE 估计联合概率密度函数 $f_{X_1, X_2}(x_1, x_2)$ 。这两节采用本书和 7 章类似的结构，方便大家比较阅读。

二元高斯分布 → 联合概率密度函数 $f_{X_1, X_2}(x_1, x_2)$

假设 (X_1, X_2) 服从二元高斯分布：

$$(X_1, X_2) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (27)$$

利用鸢尾花 150 个样本数据，我们可以估算得到 (X_1, X_2) 的质心和协方差矩阵分别为：

$$\boldsymbol{\mu} = \begin{bmatrix} 5.843 \\ 3.057 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 0.685 & -0.042 \\ -0.042 & 0.189 \end{bmatrix} \quad (28)$$

(X_1, X_2) 的联合概率密度函数 $f_{X_1, X_2}(x_1, x_2)$ 解析式则为：

$$f_{X_1, X_2}(x_1, x_2) \approx \frac{\exp \left[\frac{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{-0.739x_1^2 - 0.33x_1x_2 - 2.668x_2^2 + 9.651x_1 + 18.248x_2 - 56.093} \right]}{2\pi \times 0.358 \frac{(\sqrt{2\pi})^2}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}}} \quad (29)$$

图 21 所示为假设 (X_1, X_2) 服从二元高斯分布时，联合概率密度函数 $f_{X_1, X_2}(x_1, x_2)$ 的三维等高线和平面等高线。

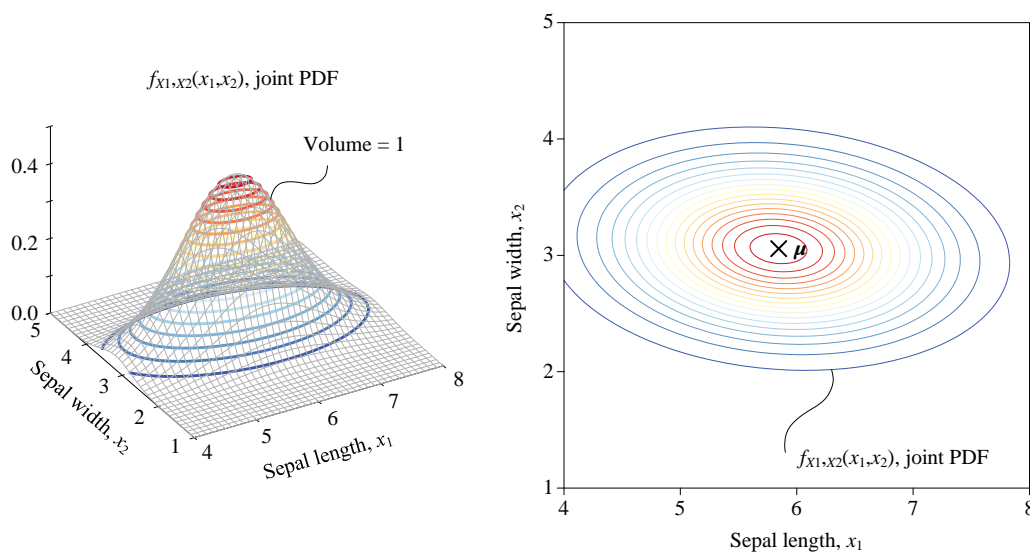


图 21. $f_{X_1, X_2}(x_1, x_2)$ 联合概率密度三维等高线和平面等高线，不考虑分类

举个例子，花萼长度 (X_1) 为 6.5、花萼宽度 (X_2) 为 2.0 时，利用 (29) 估计得到联合概率密度值为：

$$f_{X_1, X_2}(x_1 = 6.5, x_2 = 2.0) \approx 0.0205 \quad (30)$$

注意，这个数值是概率密度，不是概率。也就是说，我们不能说鸢尾花取到花萼长度 (X_1) 为 6.5、花萼宽度 (X_2) 为 2.0 时对应的概率值为 0.02097。即便这个值某种程度上也代表可能性。

马氏距离椭圆的性质

《矩阵力量》第 20 章介绍过马氏距离 (Mahalanobis distance 或 Mahal distance)，具体定义为：

$$d = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})} \quad (31)$$

图 22 所示为基于鸢尾花花萼长度、花萼宽度样本数据的马氏距离椭圆。图中，黑色旋转椭圆分别代表马氏距离为 1、2、3、4。图中，还有一个 $\mu_1 \pm \sigma_1$ 和 $\mu_2 \pm \sigma_2$ 构成的矩形。根据本章前文所学，我们知道椭圆和矩形相切于四个点。

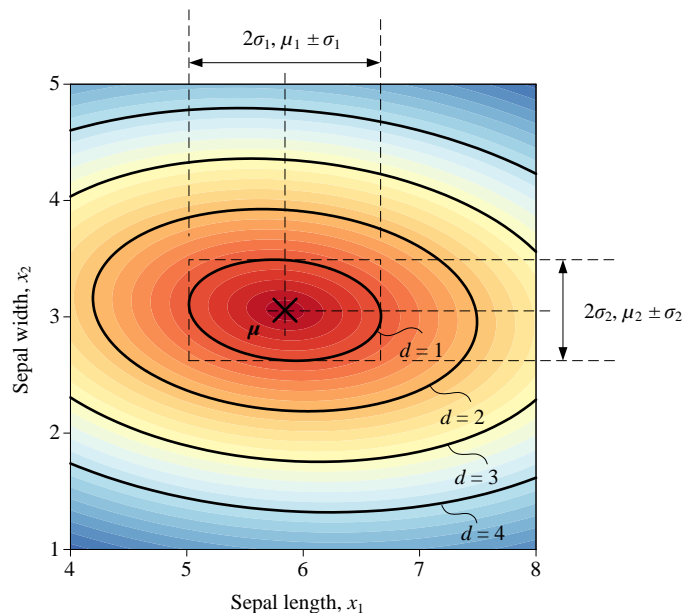


图 22. 马氏距离的椭圆，鸢尾花花萼长度、花萼宽度样本数据

还有一个需要大家注意的矩形。如图 23 所示，这个椭圆相切的“矩形”，它的长边平行于椭圆的长轴。请大家自行计算椭圆长轴倾斜角。

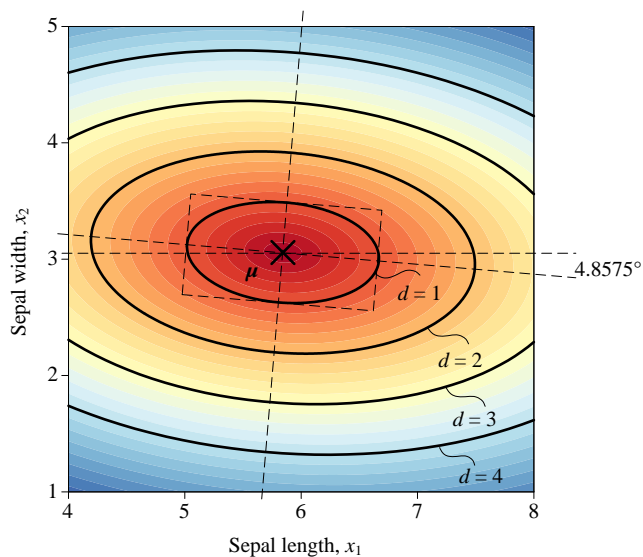


图 23. 马氏距离的椭圆的长轴、短轴，以及对应矩形

我们已经知道马氏距离和概率密度之间的关系为：

$$f_{X_1, X_2}(x_1, x_2) = \frac{\exp\left(-\frac{1}{2}d^2\right)}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} \quad (32)$$

当 $d = 1$ 时：

$$f_{X_1, X_2}(x_1, x_2)|_{d=1} = \frac{\exp\left(-\frac{1}{2} \times 1^2\right)}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} \approx 0.2693 \quad (33)$$

当 $d = 2$ 时：

$$f_{X_1, X_2}(x_1, x_2)|_{d=2} = \frac{\exp\left(-\frac{1}{2} \times 2^2\right)}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} \approx 0.0601 \quad (34)$$

如图 24 所示，利用二重积分，我们可以计算图中阴影区域 D 对应的概率：

$$\iint_D f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 \quad (35)$$

这个问题的答案留到本书第 23 章回答。

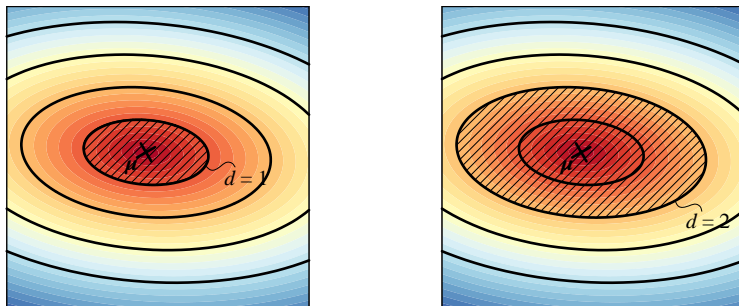
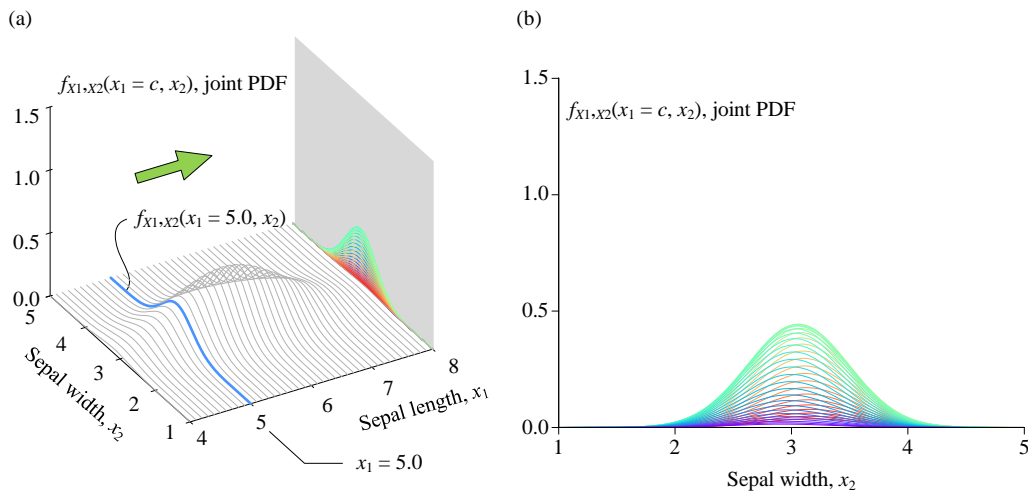
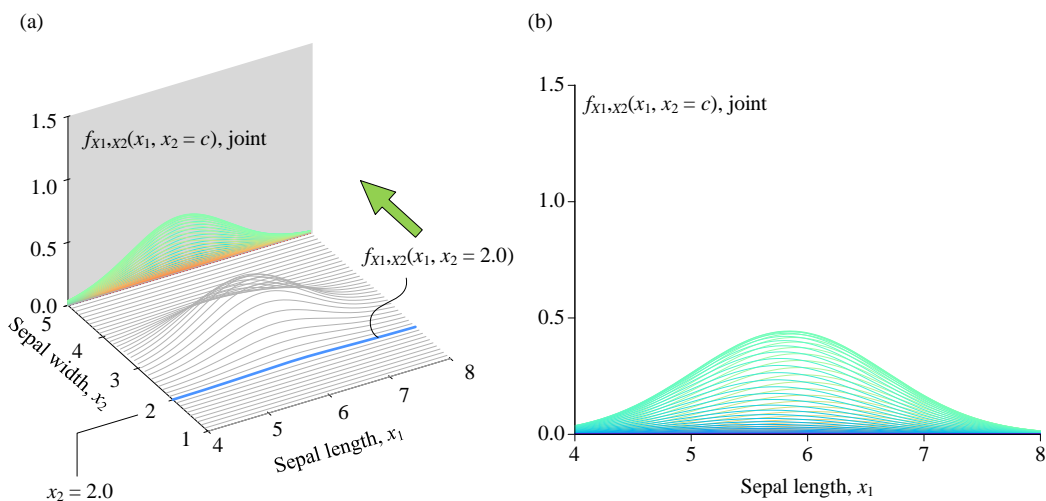


图 24. 求阴影区域对应的概率

联合概率密度函数 $f_{X_1, X_2}(x_1, x_2)$ 的剖面线

$f_{X_1, X_2}(x_1, x_2)$ 本质上是个二元函数，因此我们还可以使用“剖面线”分析二元函数。

当固定 x_1 取值时， $f_{X_1, X_2}(x_1 = c, x_2)$ 代表一条曲线。将一系列类似曲线投影到竖直平面得到图 25 (b)。观察图 25 (b)，我们容易发现这些曲线都类似一元高斯分布。图 26 所示为固定 x_2 时，概率密度函数 $f_{X_1, X_2}(x_1, x_2)$ 随 x_1 变化。

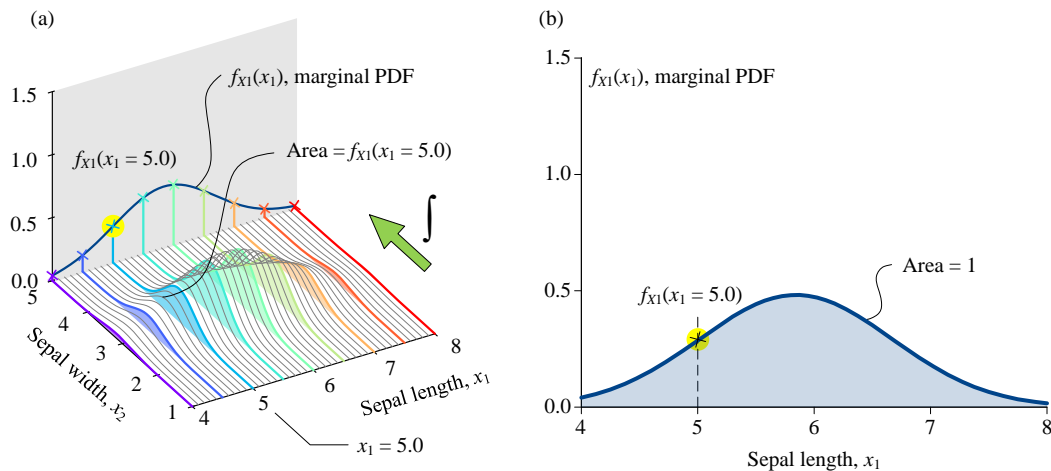
图 25. 固定 x_1 时, 概率密度函数 $f_{X1,X2}(x_1, x_2)$ 随 x_2 变化图 26. 固定 x_2 时, 概率密度函数 $f_{X1,X2}(x_1, x_2)$ 随 x_1 变化

花萼长度边缘概率密度函数 $f_{X1}(x_1)$: 偏积分

图 27 所示为求解花萼长度边缘概率 $f_{X1}(x_1)$ 的过程:

$$\underbrace{f_{X1}(x_1)}_{\text{Marginal}} = \int_{-\infty}^{+\infty} \underbrace{f_{X1,X2}(x_1, x_2)}_{\text{Joint}} dx_2 \quad (36)$$

图 27 中彩色阴影面积对应边缘概率, 即 $f_{X1}(x_1)$ 曲线高度。 $f_{X1}(x_1)$ 本身也是概率密度, 不是概率值。 $f_{X1}(x_1)$ 再积分可以得到概率。如图 27 (b) 所示, $f_{X1}(x_1)$ 曲线和整个横轴围成图形的面积为 1。本章开始时, 我们知道 $f_{X1}(x_1)$ 也是一元高斯分布 PDF。

图 27. 偏积分求解边缘概率 $f_{X1}(x_1)$

花萼长度边缘概率 $f_{X2}(x_2)$: 偏求和

图 28 所示为求解花萼宽度边缘概率的过程:

$$f_{X2}(x_2) = \int_{-\infty}^{+\infty} f_{X1,X2}(x_1, x_2) dx_1 \quad (37)$$

如所示, $f_{X2}(x_2)$ 为一元高斯分布 PDF。

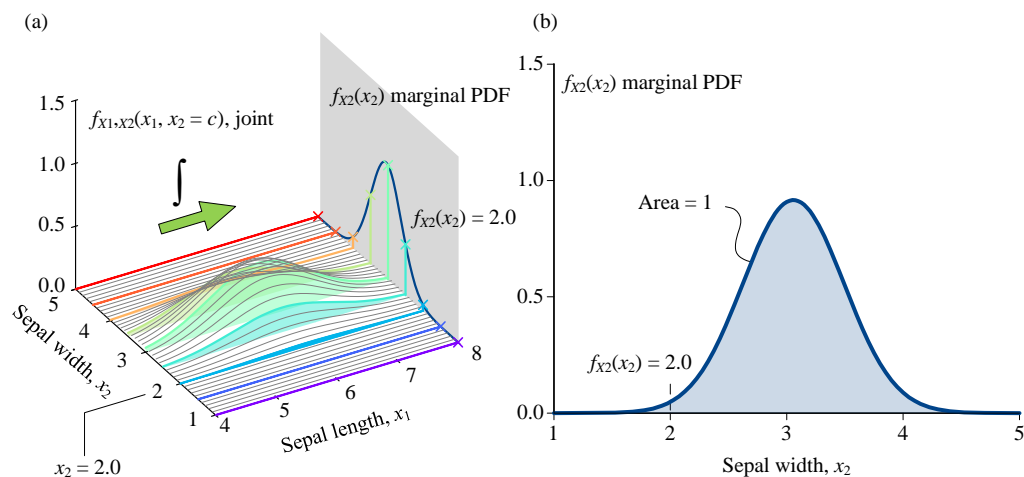
图 28. 偏积分求解边缘概率 $f_{X2}(x_2)$

图 29 所示为联合概率和边缘概率之间关系。图中联合概率密度 $f_{X1,X2}(x_1, x_2)$ 采用二元高斯分布估计得到。图 29 中 $f_{X1,X2}(x_1, x_2)$ 没有特别准确捕捉到鸢尾花花萼长度、花萼宽度样本分布细节。

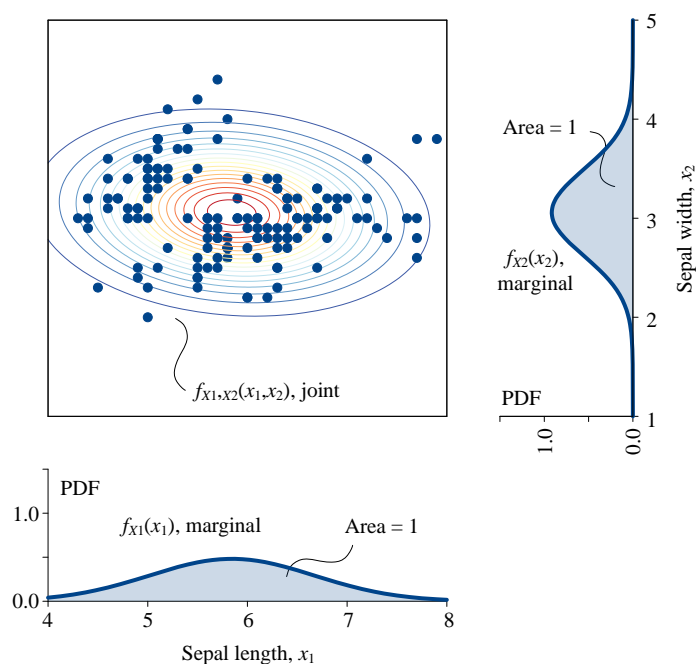


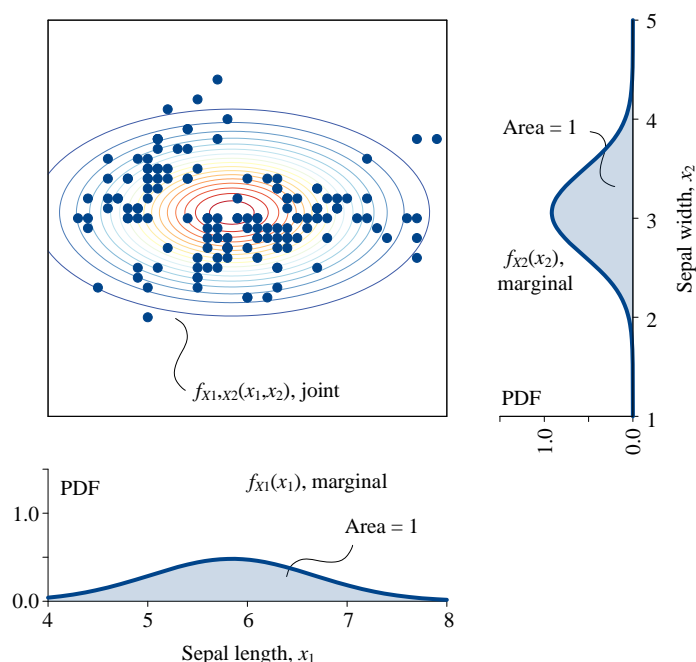
图 29. 联合概率和边缘概率之间关系

假设独立

如果假设 X_1 和 X_2 独立，则联合概率密度 $f_{X1,X2}(x1,x2)$ 可通过下式计算得到：

$$f_{X1,X2}(x1,x2) = f_{X1}(x1) \cdot f_{X2}(x2) \quad (38)$$

图 30 所示为假设 X_1 和 X_2 独立时 $f_{X1,X2}(x1,x2)$ 的平面等高线和边缘概率之间关系。椭圆等高线为正椭圆，而非旋转椭圆 (图 29)。

图 30. 联合概率，假设 X_1 和 X_2 独立

给定花萼长度，花萼宽度的条件概率密度 $f_{X_2|X_1}(x_2|x_1)$

如图 31 所示，利用贝叶斯定理，条件概率密度 $f_{X_2|X_1}(x_2|x_1)$ 可以通过下式计算：

$$\underbrace{f_{X_2|X_1}(x_2|x_1)}_{\text{Conditional}} = \frac{\overbrace{f_{X_1, X_2}(x_1, x_2)}^{\text{Joint}}}{\underbrace{f_{X_1}(x_1)}_{\text{Marginal}}} \quad (39)$$

分母中的边缘概率 $f_{X_1}(x_1)$ 起到归一化作用。如图 31 (b) 所示，经过归一化的条件概率曲线围成的面积变为 1。

将不同位置的条件概率密度 $f_{X_2|X_1}(x_2|x_1)$ 曲线投影到平面得到图 32。我们隐约发现图 32 (b) 中每条曲线看上去都是一元高斯分布。这难道是个巧合？我们将在本书第 13 章揭晓答案。

$f_{X_2|X_1}(x_2|x_1)$ 本身也是一个二元函数。图 33 所示为 $f_{X_2|X_1}(x_2|x_1)$ 三维等高线和平面等高线。从平面等高线中，我们看到一系列直线。这难道也是个巧合？答案同样在本书第 13 章给出。

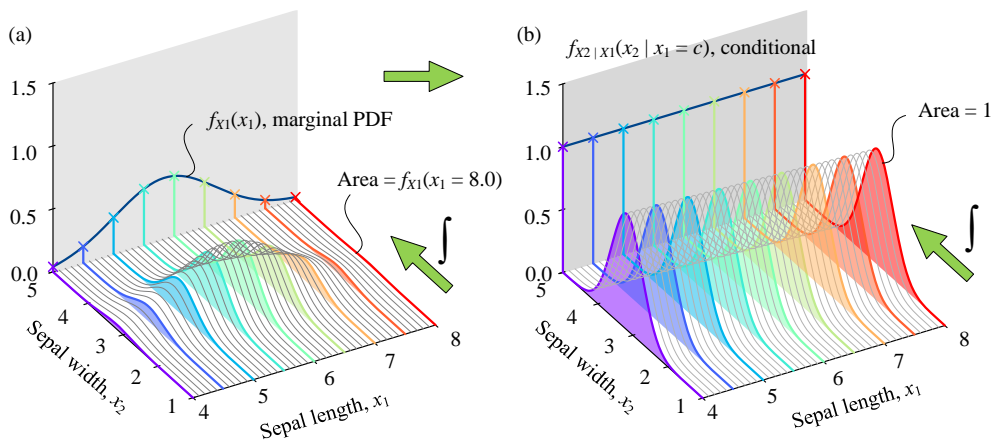


图 31. 计算条件概率 $f_{x2|x1}(x2|x1)$ 原理

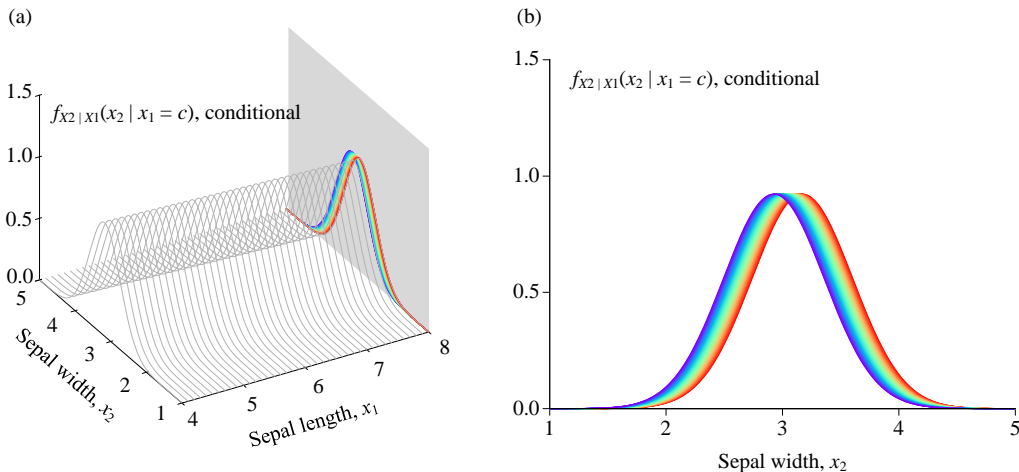


图 32. $f_{x2|x1}(x2|x1)$ 曲线投影到平面

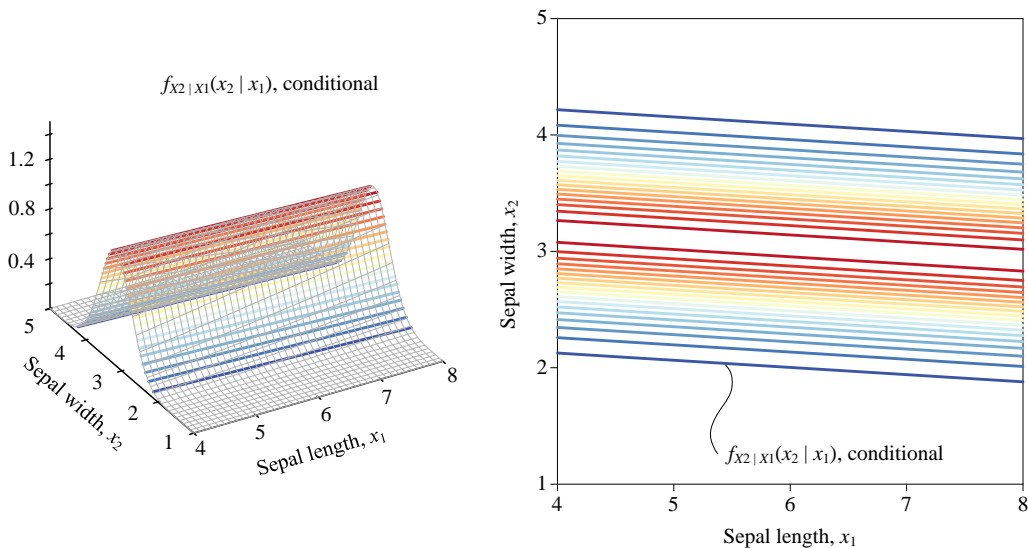


图 33. $f_{x2|x1}(x2|x1)$ 条件概率密度三维等高线和平面等高线，不考虑分类

给定花萼宽度，花萼长度的条件概率密度函数 $f_{x_1|x_2}(x_1|x_2)$

如图 34 所示，同样利用贝叶斯定理，条件概率密度 $f_{x_1|x_2}(x_1|x_2)$ 可以通过下式计算：

$$\underbrace{f_{x_1|x_2}(x_1|x_2)}_{\text{Conditional}} = \frac{\overbrace{f_{x_1,x_2}(x_1,x_2)}^{\text{Joint}}}{\underbrace{f_{x_2}(x_2)}_{\text{Marginal}}} \quad (40)$$

类似前文，上式中分母中 $f_{x_2}(x_2)$ 起到归一化作用。

将不同位置的条件概率密度 $f_{x_1|x_2}(x_1|x_2)$ 曲线投影到平面得到图 35。图 35 (b) 中每条曲线也都像是一元高斯分布曲线。

$f_{x_1|x_2}(x_1|x_2)$ 同样也是一个二元函数，如图 36 所示的 $f_{x_1|x_2}(x_1|x_2)$ 三维等高线和平面等高线。

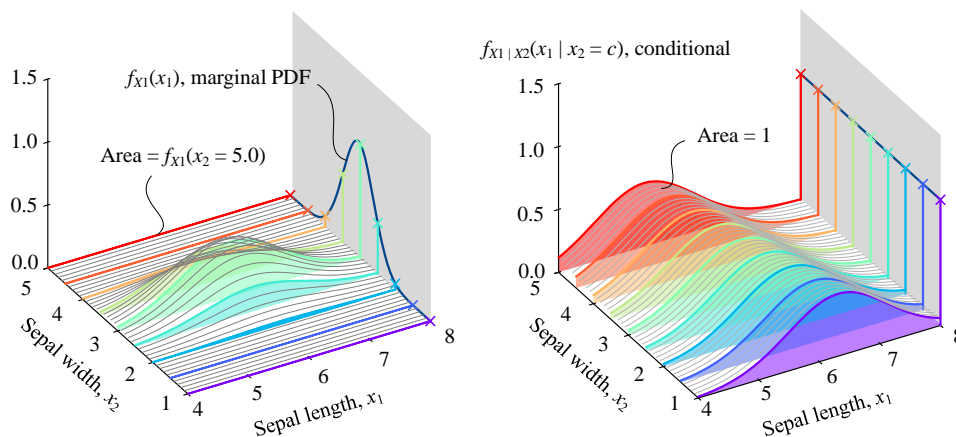


图 34. 计算条件概率 $f_{x_1|x_2}(x_1|x_2)$ 原理

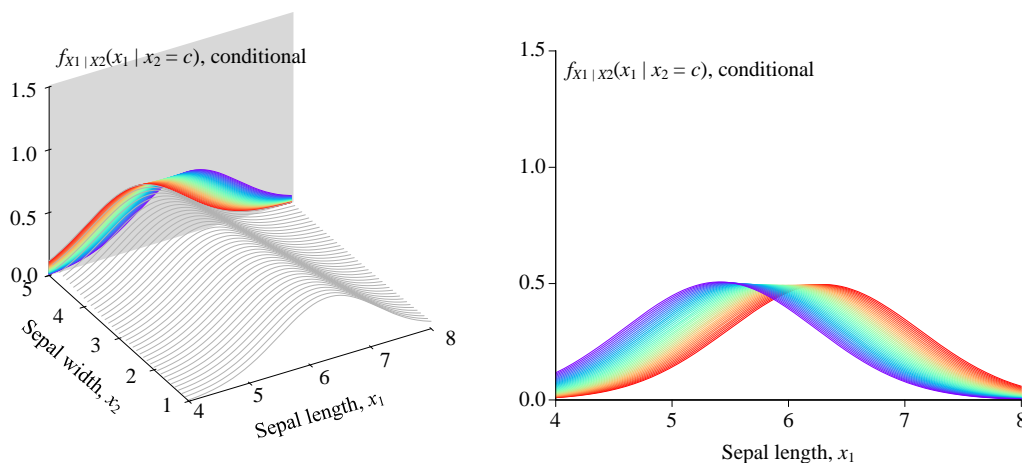
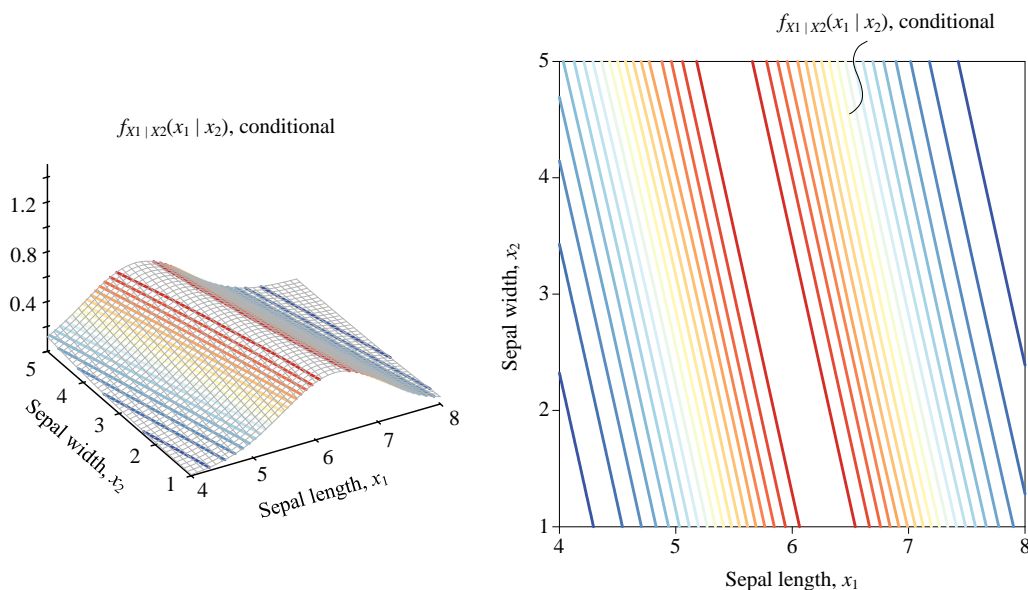


图 35. $f_{x_1|x_2}(x_1|x_2)$ 曲线投影到平面

图 36. $f_{X1|X2}(x1|x2)$ 条件概率密度三维等高线和平面等高线，不考虑分类

10.6 以鸢尾花数据为例：考虑分类标签

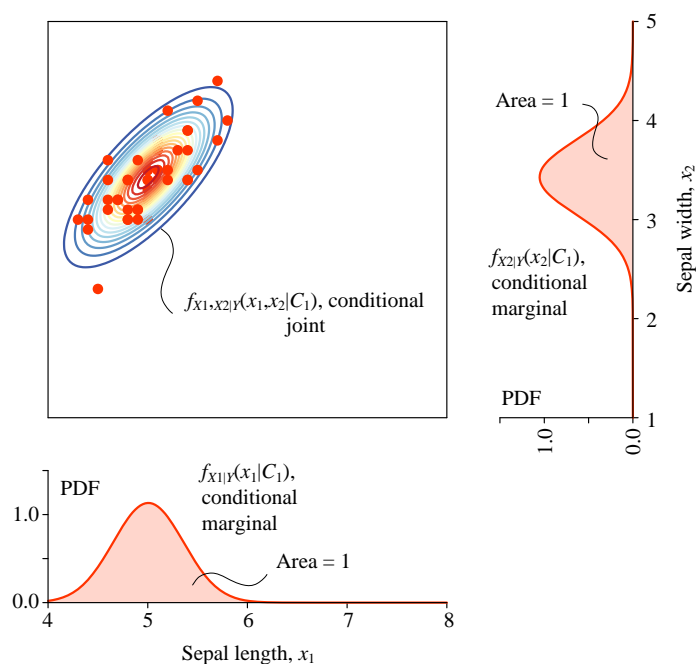
本节讨论考虑鸢尾花分类条件下的条件概率 PDF。

给定分类标签 $Y = C_1$ (setosa)

给定分类标签 $Y = C_1$ (setosa) 条件下，假设鸢尾花花萼长度、花萼宽度同样服从二元高斯分布。

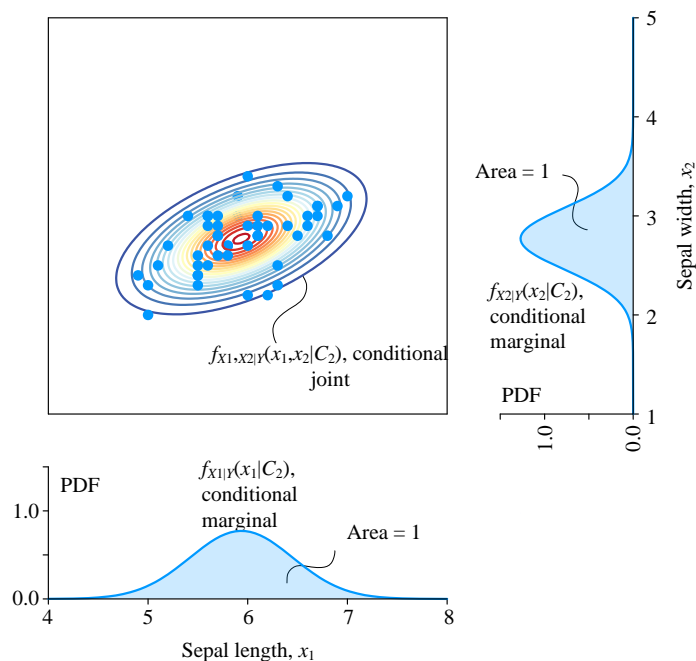
图 37 所示为，给定分类标签 $Y = C_1$ (setosa)，条件概率 $f_{X1,X2|Y}(x1, x2 | y = C1)$ 平面等高线和条件边缘概率密度曲线。 $f_{X1,X2|Y}(x1, x2 | y = C1)$ 曲面和整个水平面围成体积为 1。

图 37 中 $f_{X1|Y}(x1 | y = C1)$ 、 $f_{X2|Y}(x2 | y = C1)$ 分别和 $x1$ 、 $x2$ 围成的面积也是 1。

图 37. 条件概率 $f_{X1, X2|Y}(x1, x2 | y = C1)$ 平面等高线和条件边缘概率密度曲线，给定分类标签 $Y = C1$ (setosa)

给定分类标签 $Y = C2$ (versicolor)

图 38 所示为，给定分类标签 $Y = C2$ (versicolor)，条件概率 $f_{X1, X2|Y}(x1, x2 | y = C2)$ 平面等高线和条件边缘概率密度曲线。

图 38. 条件概率 $f_{X1, X2|Y}(x1, x2 | y = C2)$ 平面等高线和条件边缘概率密度曲线，给定分类标签 $Y = C2$ (versicolor)

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

给定分类标签 $Y = C_3$ (virginica)

图 39 所示为，给定分类标签 $Y = C_3$ (virginica)，条件概率 $f_{X_1, X_2 | Y}(x_1, x_2 | y = C_3)$ 平面等高线和条件边缘概率密度曲线。

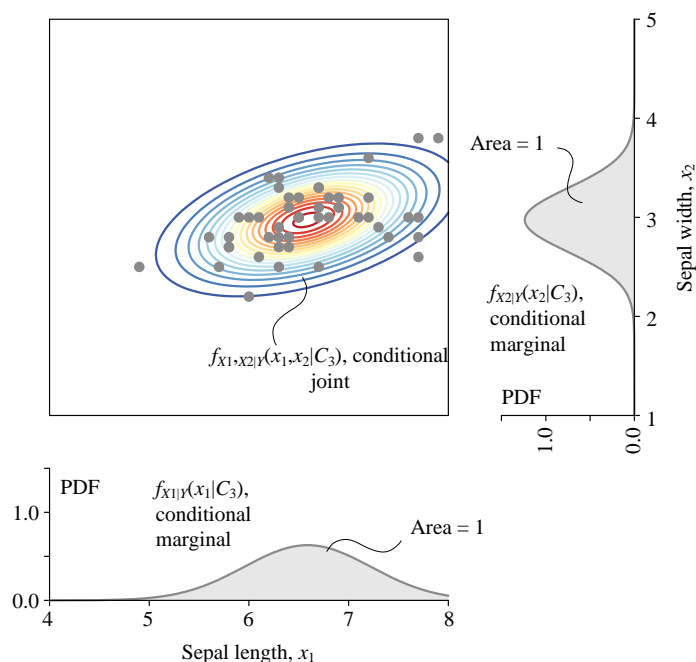


图 39. 条件概率 $p_{X_1, X_2 | Y}(x_1, x_2 | y = C_3)$ 平面等高线和条件边缘概率密度曲线，给定分类标签 $Y = C_3$ (virginica)

全概率

如图 40 所示，利用全概率定理，利用全概率定理，三幅条件概率等高线叠加可以得到联合概率密度，即：

$$\begin{aligned}
 f_{X_1, X_2}(x_1, x_2) &= f_{X_1, X_2 | Y}(x_1, x_2 | y = C_1) p_Y(C_1) + \\
 &\quad f_{X_1, X_2 | Y}(x_1, x_2 | y = C_2) p_Y(C_2) + \\
 &\quad f_{X_1, X_2 | Y}(x_1, x_2 | y = C_3) p_Y(C_3)
 \end{aligned} \tag{41}$$

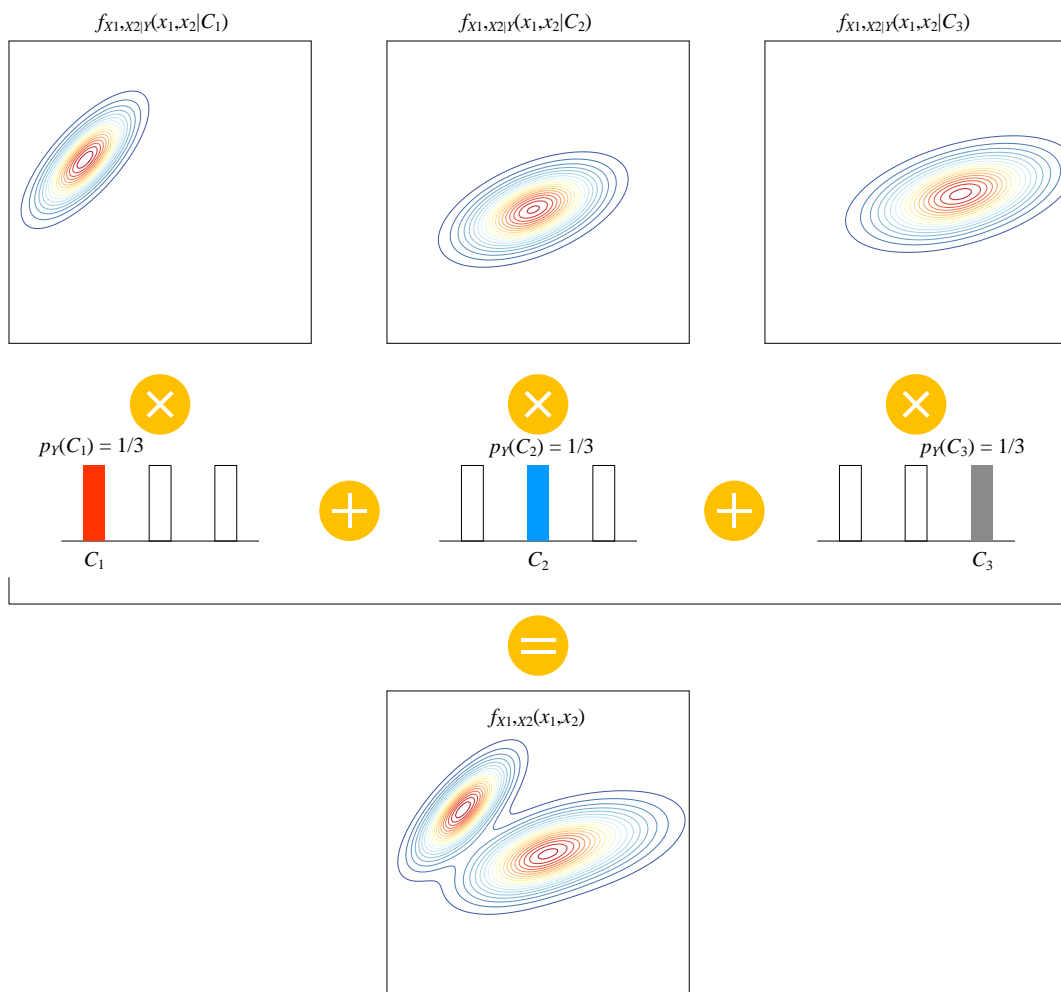


图 40. 估算联合概率密度，假设条件概率服从二元高斯分布

假设条件独立

如图 41 所示，如果假设条件独立， $f_{X1,X2|Y}(x1,x2|y=C1)$ 可以通过下式计算得到：

$$\underbrace{f_{X1,X2|Y}(x1,x2|y=C1)}_{\text{Conditional joint}} = \underbrace{f_{X1|Y}(x1|y=C1)}_{\text{Conditional marginal}} \cdot \underbrace{f_{X2|Y}(x2|y=C1)}_{\text{Conditional marginal}} \quad (42)$$

同理我们可以计算得到 $f_{X1,X2|Y}(x1,x2|y=C2)$ 、 $f_{X1,X2|Y}(x1,x2|y=C3)$ ，具体如图 42、图 43 所示。

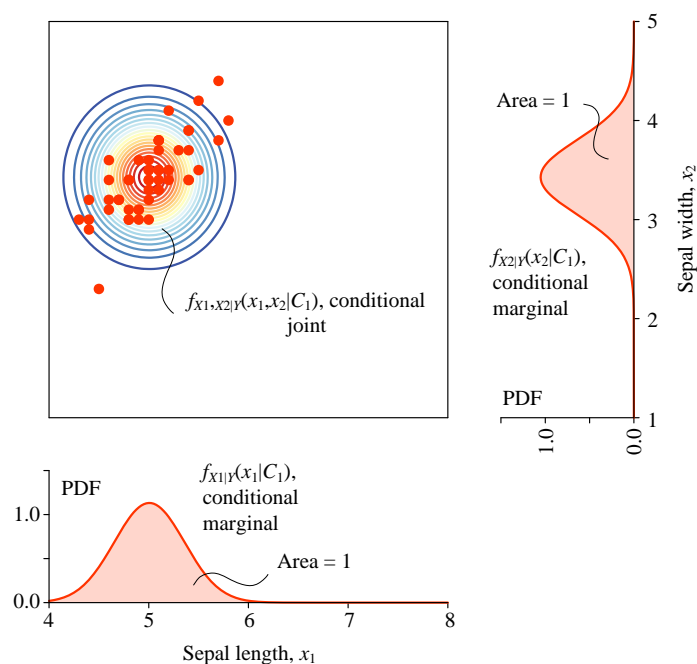


图 41. 给定 $Y = C_1$, X_1 和 X_2 条件独立, 估算条件概率 $f_{X1, X2|Y}(x_1, x_2 | y = C_1)$

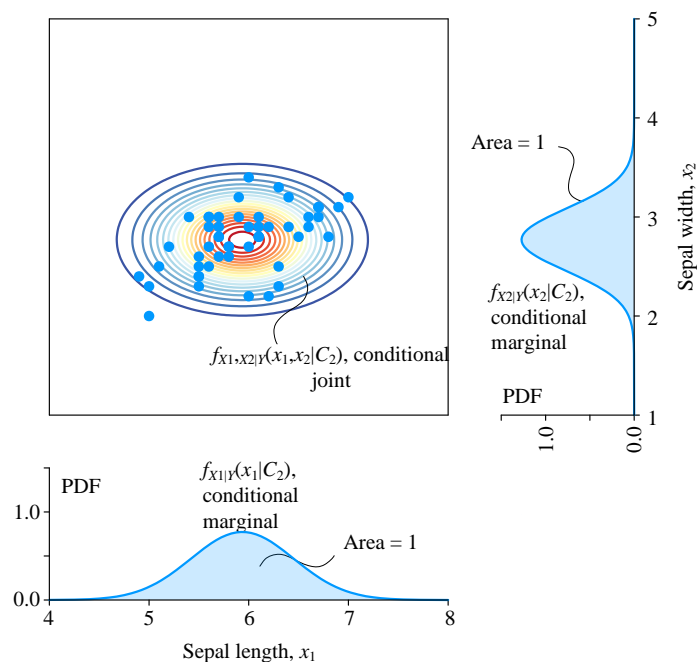
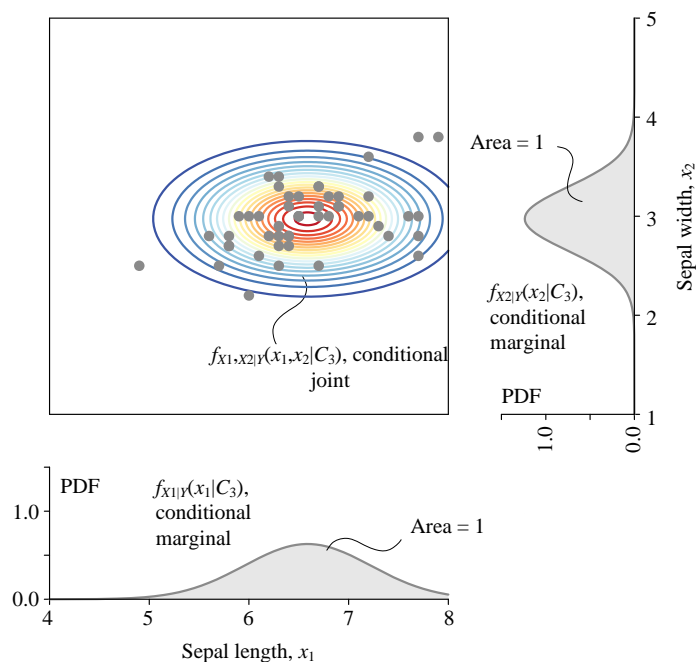


图 42. 给定 $Y = C_2$, X_1 和 X_2 条件独立, 估算条件概率 $f_{X1, X2|Y}(x_1, x_2 | y = C_2)$

图 43. 给定 $Y = C_3$, X_1 和 X_2 条件独立, 估算条件概率 $f_{X1,X2|Y}(x_1, x_2 | y = C_3)$

估计联合概率

如图 44 所示, 并利用全概率定理, 估算 $f_{X1,X2}(x_1, x_2)$:

$$\begin{aligned}
 f_{X1,X2}(x_1, x_2) &= f_{X1|Y}(x_1 | y = C_1) f_{X2|Y}(x_2 | y = C_1) p_Y(C_1) + \\
 &\quad f_{X1|Y}(x_1 | y = C_2) f_{X2|Y}(x_2 | y = C_2) p_Y(C_2) + \\
 &\quad f_{X1|Y}(x_1 | y = C_3) f_{X2|Y}(x_2 | y = C_3) p_Y(C_3) +
 \end{aligned} \tag{43}$$

图 40 和图 44 涉及的这些技术细节对于理解贝叶斯分类器原理有很重要意义。本书第 19、20 章将从贝叶斯定理视角简单介绍分类原理, 《机器学习》一册将专门讲解朴素贝叶斯分类器。

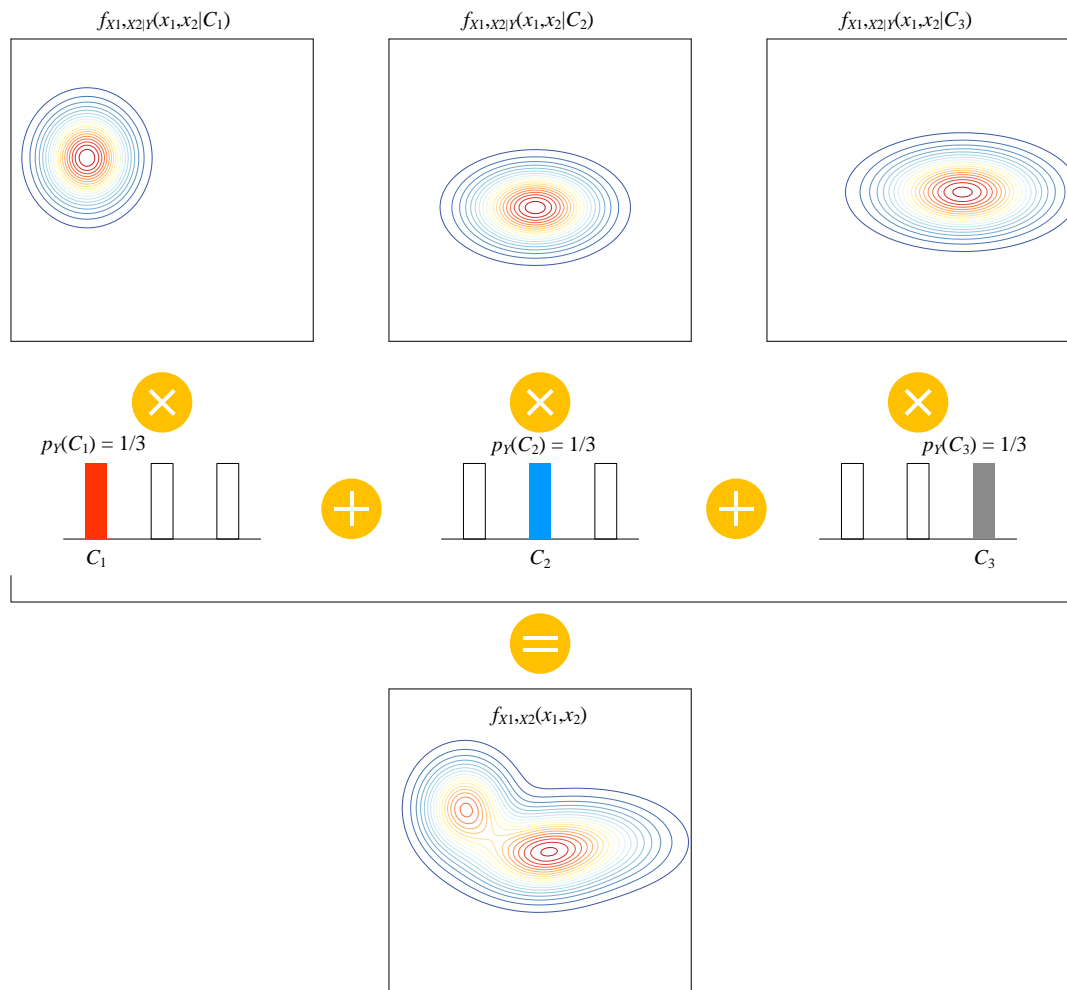


图 44. 利用全概率定理，估算 $f_{X1,X2}(x1, x2)$ ，假设条件独立

