

12

Conditional Gaussian Distributions

条件高斯分布

假设随机变量服从高斯分布，讨论条件期望、条件方差



生命就像一个永恒的春天，穿着崭新而绚丽的衣服站在我面前。

Life stands before me like an eternal spring with new and brilliant clothes.

—— 卡尔·弗里德里希·高斯 (Carl Friedrich Gauss) | 德国数学家、物理学家、天文学家 | 1777 ~ 1855



- ▶ `matplotlib.pyplot.contour()` 绘制等高线图
- ▶ `matplotlib.pyplot.contour3D()` 绘制三维等高线图
- ▶ `matplotlib.pyplot.contourf()` 绘制填充等高线图
- ▶ `matplotlib.pyplot.fill_between()` 区域填充颜色
- ▶ `matplotlib.pyplot.plot_wireframe()` 绘制线框图
- ▶ `scipy.stats.multivariate_normal()` 多元正态分布对象
- ▶ `scipy.stats.norm()` 一元正态分布对象

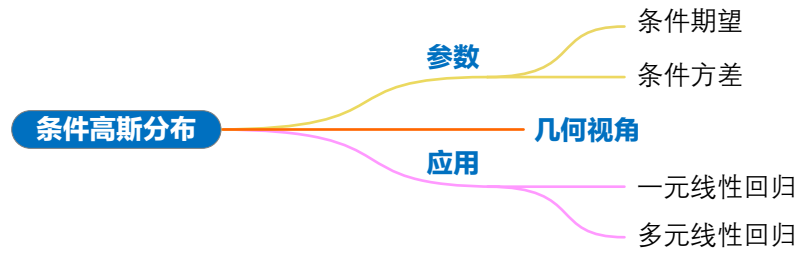
本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

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



12.1 联合概率和条件概率关系

本章是本书第 8 章的延续。本书第 8 章专门介绍了离散、连续随机变量的条件期望 (conditional expectation)、条件方差 (conditional variance)。本章将这些数学工具用在高斯分布上。

本节首先回顾条件概率 (conditional probability)。

条件概率

本章第 3 章介绍过，条件概率是指某事件在另外一个事件已经发生条件下的概率。

以图 1 为例， X 和 Y 为连续随机变量， (X, Y) 服从二元高斯分布。 (X, Y) 的联合概率密度函数 PDF $f_{X,Y}(x,y)$ 为图 1 所示曲面。

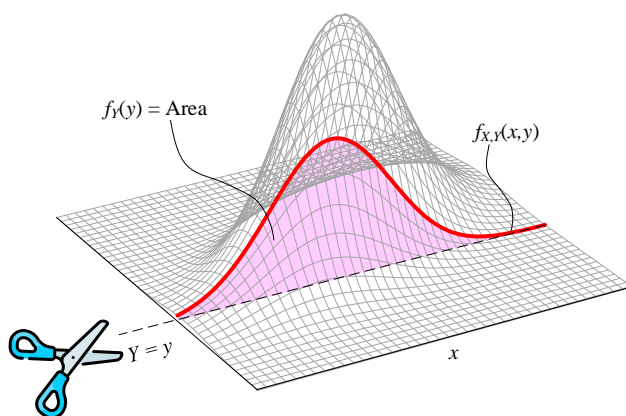


图 1. 高斯二元分布 PDF 曲面沿着 $Y=y$ 切一刀

给定 $Y=y$ 条件下，相当于在图 1 上沿着 $Y=y$ 切一刀，得到的红色曲线便是 $f_{X|Y}(x, y)$ 。

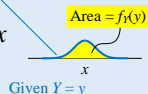
几何视角来看，给定 $Y=y$ 的条件下 ($f_Y(y) > 0$)，利用贝叶斯定理， X 的条件 PDF $f_{X|Y}(x|y)$ 相当于对 $f_{X,Y}(x, y)$ 曲线用边缘 PDF $f_Y(y)$ 归一化：

$$f_{X|Y}(x|y) = \frac{\text{Joint } f_{X,Y}(x,y)}{\text{Marginal } f_Y(y)} \quad (1)$$

Diagram illustrating the relationship between the joint PDF, conditional PDF, and marginal PDF. The joint PDF is shown as a 3D surface. The conditional PDF is shown as a 2D curve on the surface. The marginal PDF is shown as a 2D curve. The area under the joint PDF is labeled Area = 1. The area under the conditional PDF is labeled Area = f_Y(y). The marginal PDF is labeled f_Y(y).

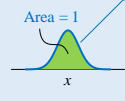
⚠ 注意，此时 $f_Y(y)$ 代表一个具体的值，但是这个值仍然是概率密度，而不是概率。

分解来看， $Y=y$ 时，联合 PDF $f_{X,Y}(x, y)$ 这条曲线和横轴围成的面积为边缘 PDF $f_Y(y)$ ，即：

$$\underbrace{f_Y(y)}_{\text{Marginal}} = \int_x \underbrace{f_{X,Y}(x,y)}_{\text{Joint}} dx$$


(2)

归一化后的 $f_{X|Y}(x|y)$ 曲线和横轴围成的面积为 1，即：



$$\int_x \underbrace{f_{X|Y}(x|y)}_{\text{Conditional}} dx = 1$$

(3)

沿着这个思路，让我们观察一组当 Y 取不同值时，高斯二元分布联合概率和条件概率的关系。

Y 取特定值

如图 2 所示，当 $y = -2$ 时，在联合 PDF 曲面在 $y = -2$ 处切一刀，得到 $f_{X,Y}(x, y = -2)$ 对应图 2 中红色曲线。

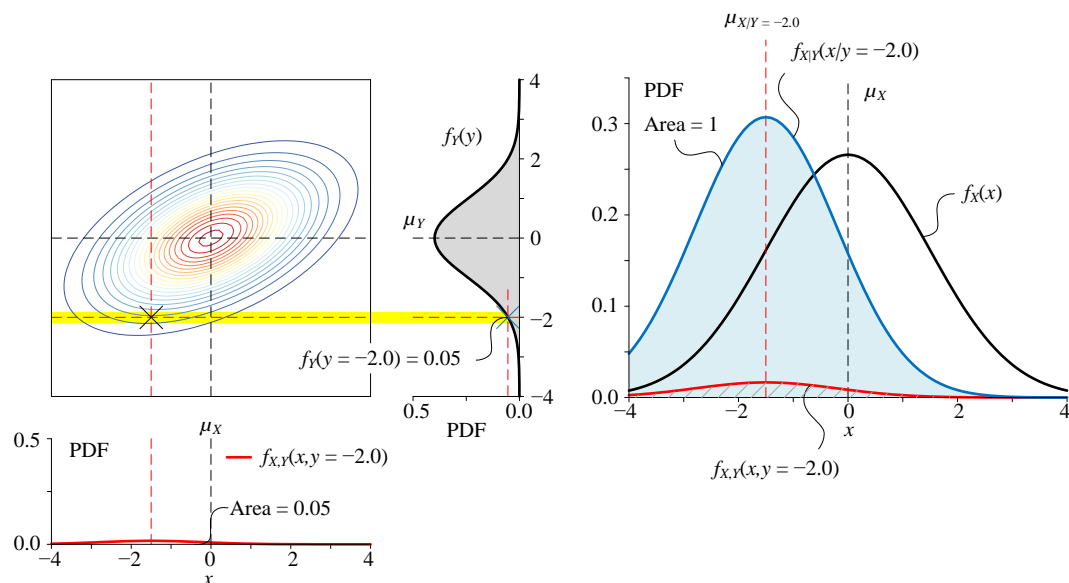
$f_{X,Y}(x, y = -2)$ 和横轴围成的面积便是边缘 PDF $f_Y(y = -2)$ ，经过计算得知面积约为 0.05，即 $f_Y(y = -2) = 0.05$ 。

再次强调，0.05 不是概率值，虽然它的大小某种程度上也代表“可能性”。

在给定 $y = -2$ 条件下，条件 PDF $f_{X|Y}(x|y = -2)$ 可以通过下式计算得到：

$$f_{X|Y}(x|y = -2) = \frac{f_{X,Y}(x, y = -2)}{f_Y(y = -2)} \quad (4)$$

图 2 同时比较联合 PDF $f_{X,Y}(x, y = -2)$ 、边缘 PDF $f_X(x)$ 、条件 PDF $f_{X|Y}(x|y = -2)$ 三条曲线之间的关系。

图 2. $y = -2$ 时，联合 PDF、边缘 PDF、条件 PDF 的关系

从图像上可以清楚看到，条件 PDF $f_{X|Y}(x|y = -2)$ 相当于联合 PDF $f_{X,Y}(x, y = -2)$ 在高度上放大大约 20 倍 ($= 1/0.05$)。

▲ 值得反复强调的是，联合 PDF $f_{X,Y}(x, y = -2)$ 曲线和横轴围成的面积约为 0.05，然而条件 PDF $f_{X|Y}(x|y = -2)$ 曲线和横轴围成的面积为 1。

Y 取不同值

图 2 到图 6 五幅图分别展示当 y 取值分别为 -2、-1、0、1、2 时，联合 PDF 和条件 PDF 关系。

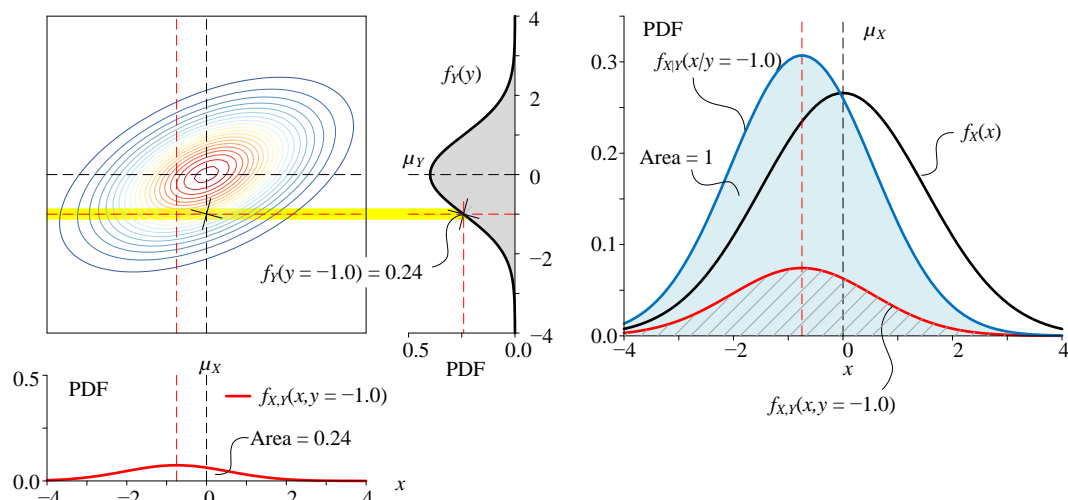
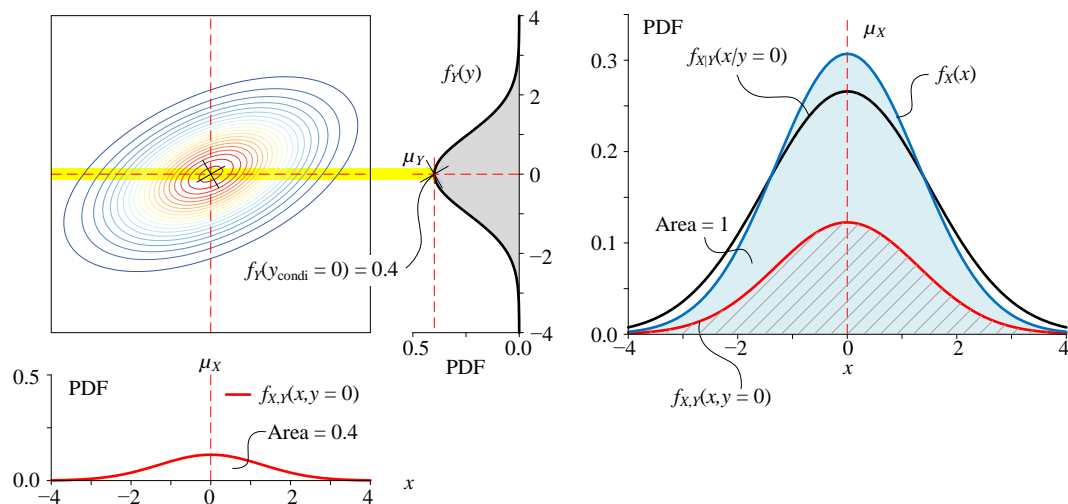
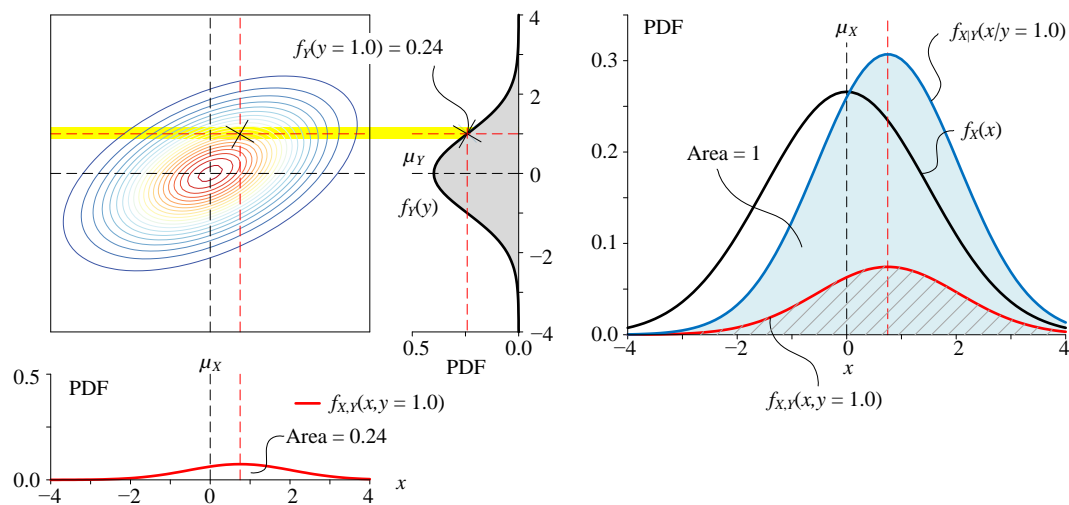
有几点值得注意。五幅图像上概率曲线形状都是类似高斯一元分布曲线。

$Y = y$ 直线和联合 PDF 等高线某一个椭圆椭圆相切，而当 y 变化时，切点似乎沿着直线运动。

切点的横轴取值对应条件 PDF $f_{X|Y}(x|y)$ 曲线的对称轴，而这个对称轴又是条件 PDF $f_{X|Y}(x|y)$ 曲线的期望。这个期望值就是本书第 8 章介绍的条件期望 (conditional expectation) $E(X|Y = y)$ 。

图 2 到图 6 五幅图条件 PDF $f_{X|Y}(x|y)$ 对应的蓝色曲线，似乎在形状上没有任何变化，仅仅是对称轴发生移动。这一点说明， y 取值变化时，条件 PDF 曲线对应分布的方差似乎没有变化；这个方差就是本书第 8 章介绍的条件方差 (conditional variance) $\text{var}(X|Y = y)$ 。

这一节先给大家一个直观印象，本章之后将会利用高斯二元分布对条件概率、条件期望、条件方差等概念进行定量研究。

图 3. $y = -1$ 时，联合 PDF、边缘 PDF、条件 PDF 的关系图 4. $y = 0$ 时，联合 PDF、边缘 PDF、条件 PDF 的关系图 5. $y = 1$ 时，联合 PDF、边缘 PDF、条件 PDF 的关系

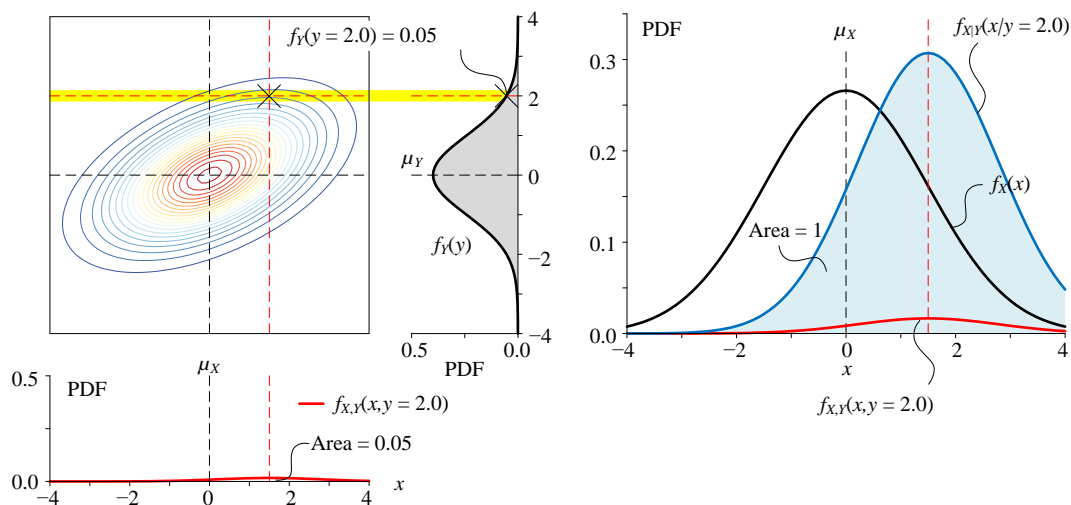
本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

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 6. $y = 2$ 时，联合 PDF、边缘 PDF、条件 PDF 的关系

Bk5_Ch12_01.py 绘制图 2 ~ 图 6。

12.2 给定 X 条件下， Y 的条件概率：以二元高斯为例

如果 (X, Y) 服从二元高斯分布，联合 PDF $f_{X,Y}(x,y)$ 解析式如下：

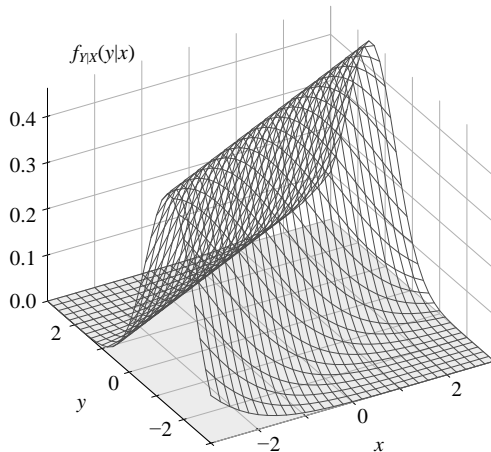
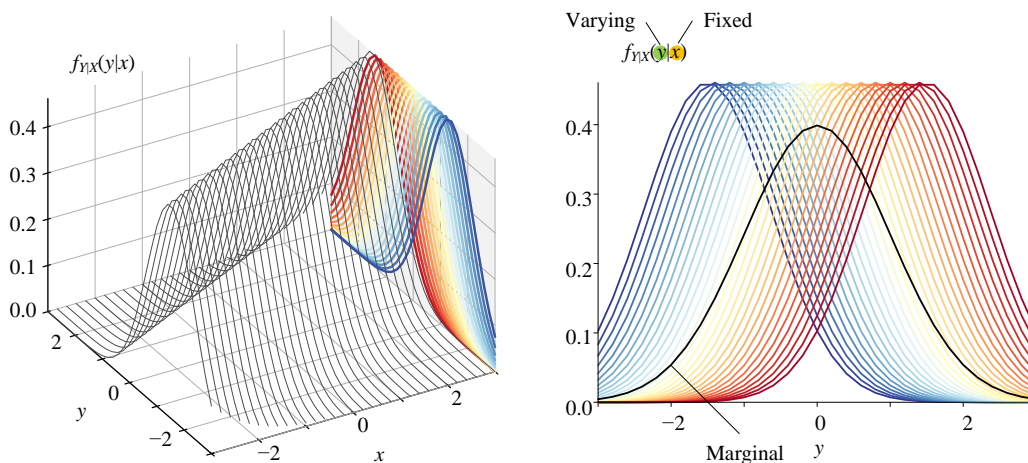
$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \times \exp \left(\underbrace{-\frac{1}{2(1-\rho_{X,Y}^2)} \left(\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho_{X,Y} \left(\frac{x-\mu_X}{\sigma_X} \right) \left(\frac{y-\mu_Y}{\sigma_Y} \right) + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right)}_{\text{Ellipse}} \right) \quad (5)$$

利用条件 PDF、联合 PDF、边缘 PDF 三者关系，我们可以求得在给定 $X = x$ 条件下，条件 PDF $f_{Y|X}(y|x)$ 解析式为：

$$f_{Y|X}(y|x) = \frac{1}{\sigma_Y\sqrt{1-\rho_{X,Y}^2}\sqrt{2\pi}} \exp \left(-\frac{1}{2} \left(\frac{y - \left(\mu_Y + \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} (x - \mu_X) \right)}{\sigma_Y\sqrt{1-\rho_{X,Y}^2}} \right)^2 \right) \quad (6)$$

图 7 所示为 $f_{Y|X}(y|x)$ 曲面网格线。 $f_{Y|X}(y|x)$ 这个曲线的期望和方差对应条件期望 $E(Y|X=x)$ 和条件方差 $\text{var}(Y|X=x)$ 。

可以发现当 $X = x$ 取一定值时，(6) 解析式对应高斯正态分布，这印证了本书第 10 章的猜测。将 $f_{Y|X}(y|x)$ 曲面不同位置曲线投影在 yz 平面得到图 8，容易发现这些曲线的形状完全相同（也就是条件**标准差**不变），但是曲线的中心位置变化（也就是条件**期望**值变化）。

图 7. $f_{Y|X}(y|x)$ 曲面网格线图 8. $f_{Y|X}(y|x)$ 曲面在 yz 平面上投影

条件期望 $E(Y|X = x)$

如果 (X, Y) 满足二元高斯分布，给定 $X = x$ 条件下， Y 的条件 PDF $f_{Y|X}(y|x)$ 如图 9 所示。图 10 所示为 $f_{Y|X}(y|x)$ 平面等高线。条件期望 $E(Y|X = x)$ 解析式为：

$$E(Y|X = x) = \mu_Y + \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} (x - \mu_X) \quad (7)$$

如图 10 所示， $E(Y|X = x)$ 随着 $X = x$ 取值线性变化；也就是说， $E(Y|X = x)$ 和 x 的关系是一条直线。这条直线的一般式可以写成：

$$y = \mu_Y + \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} (x - \mu_X) \quad (8)$$

可以发现它的斜率为 $\rho_{X,Y}\sigma_Y/\sigma_X$ ，且通过点 (μ_X, μ_Y) 。眼尖的读者一眼就会发现，这条曲线是 x 自变量、 y 为因变量的 OLS 线性回归曲线解析式。



本章最后一节将深入探讨这一话题，此外本书第 24 章也会展开讲解线性回归。

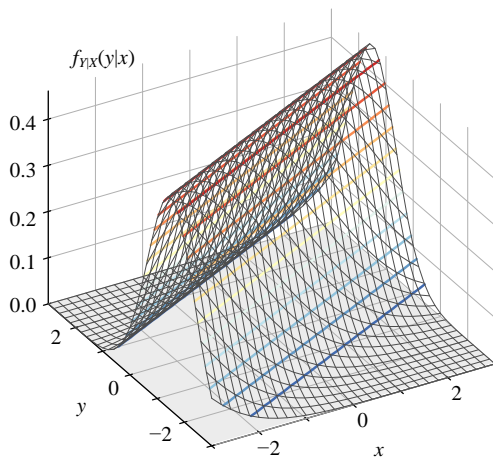


图 9. $f_{Y|X}(y|x)$ 曲面等高线

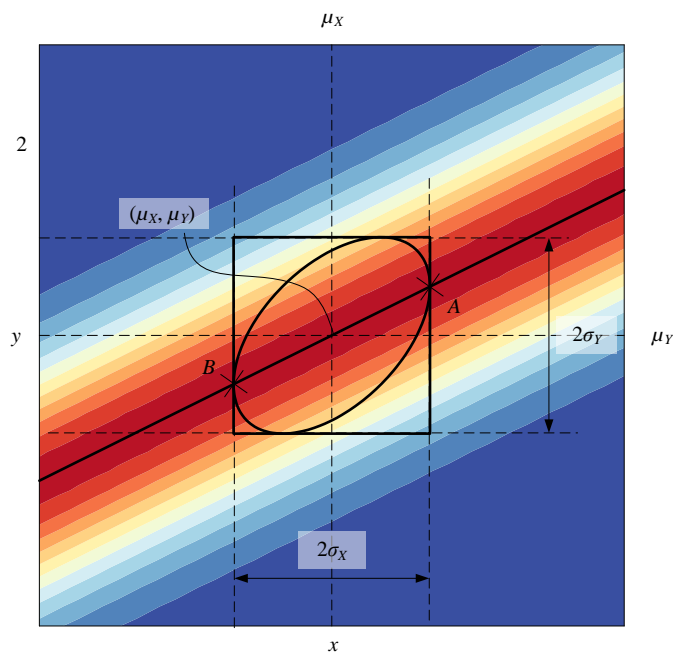


图 10. $f_{Y|X}(y|x)$ 平面等高线

条件方差 $\text{var}(Y|X=x)$

给定 $X=x$ 条件下， Y 的条件方差 $\text{var}(Y|X=x)$ 解析式为：

$$\text{var}(Y|X=x) = (1 - \rho_{X,Y}^2) \sigma_Y^2 \quad (9)$$

给定 $X=x$ 条件下, Y 的条件**标准差** $\sigma_{Y|X=x}$ 解析式为定值:

$$\sigma_{Y|X=x} = \sqrt{1 - \rho_{X,Y}^2} \cdot \sigma_Y \quad (10)$$

这解释了为什么图 10 中的等高线为平行线。

图 11 所示为 $\sigma_{Y|X=x}$ 的几何含义。



请大家格外注意图中的平行四边形, 我们将在本书第 15 章还会看到这个平行四边形。

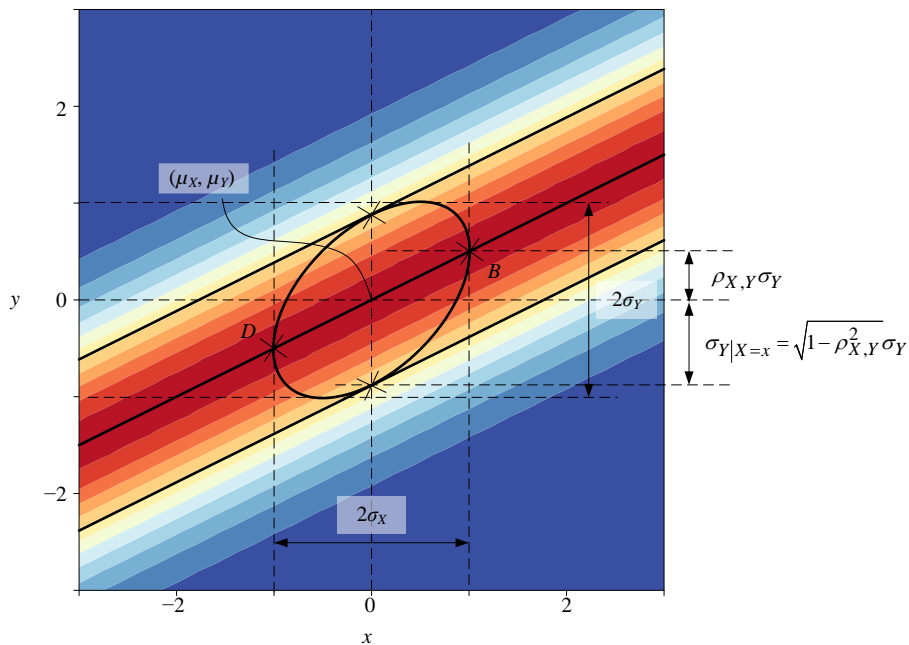
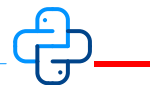


图 11. 条件**标准差** $\sigma_{Y|X}$ 的几何含义



Bk5_Ch12_02.py 绘制图 7 ~ 图 10。

以鸢尾花为例: 条件**期望** $E(X_2 | X_1 = x_1)$ 、条件**方差** $\text{var}(X_2 | X_1 = x_1)$

以鸢尾花花萼长度 (X_1)、花萼宽度 (X_2) 数据为例, 假设 (X_1, X_2) 服从二元高斯分布。条件 PDF $f_{X_2|X_1}(x_2 | x_1)$ 三维等高线和平面等高线如图 12 所示。

在给定 $X_1 = x_1$ 条件下, X_2 的条件**期望** $E(X_2 | X_1 = x_1)$ 解析式为:

$$\begin{aligned}
 E(X_2 | X_1 = x_1) &= \mu_2 + \rho_{1,2} \frac{\sigma_2}{\sigma_1} (x_1 - \mu_1) \\
 &= 3.057 - 0.117 \times \frac{0.434}{0.825} (x_1 - 5.843) \\
 &= -0.615x_1 + 3.417
 \end{aligned} \tag{11}$$

条件**方差** $\text{var}(X_2 | X_1 = x_1)$ 为：

$$\text{var}(X_2 | X_1 = x_1) = (1 - \rho_{1,2}^2) \sigma_2^2 \approx 0.186 \tag{12}$$

条件**标准差** $\sigma_{X_2 | X_1 = x_1}$ 为：

$$\sigma_{X_2 | X_1 = x_1} = \sqrt{1 - \rho_{1,2}^2} \sigma_2 = 0.431 \tag{13}$$

如图 12 所示，不管 x_1 怎么变，这个条件**标准差** $\sigma_{X_2 | X_1 = x_1}$ 为定值。请大家对比第 8 章的类似图片。

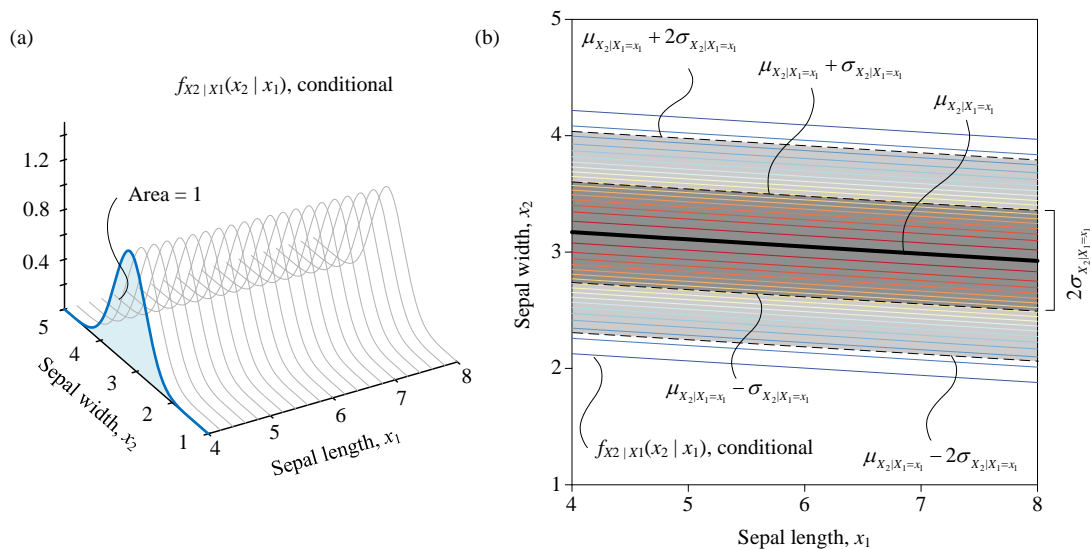
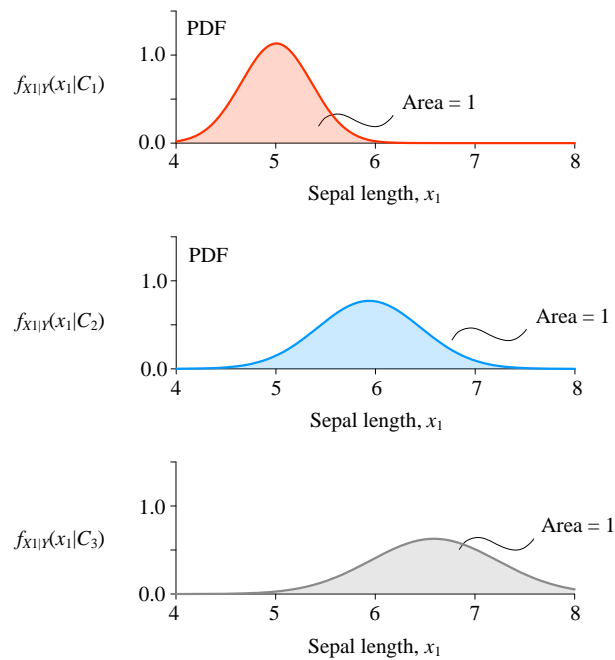


图 12. 条件 PDF $f_{X2|X1}(x_2 | x_1)$ 三维等高线和平面等高线，不考虑分类

以鸢尾花为例，考虑标签

换个条件来看，如图 13 所示，给定鸢尾花分类条件，假设花萼长度服从高斯分布。请大家自行计算给定鸢尾花分类为条件，花萼长度的条件**期望** $E(X_1 | Y = C_k)$ 、条件**方差** $\text{var}(X_1 | Y = C_k)$ 。

图 13. 给定鸢尾花标签 Y , 花萼长度的 PDF, 连续随机变量

12.3 给定 Y 条件下, X 的条件概率: 以二元高斯为例

如果 (X, Y) 服从二元高斯分布, 给定 $Y = y$ 条件下, X 的条件 PDF $f_{X|Y}(x|y)$ 解析式为:

$$f_{X|Y}(x|y) = \frac{1}{\sigma_X \sqrt{1 - \rho_{X,Y}^2} \sqrt{2\pi}} \exp \left(-\frac{1}{2} \left(\frac{x - \left(\mu_X + \rho_{X,Y} \frac{\sigma_X}{\sigma_Y} (y - \mu_Y) \right)}{\sigma_X \sqrt{1 - \rho_{X,Y}^2}} \right)^2 \right) \quad (14)$$

图 14 所示为 $f_{X|Y}(x|y)$ 网格线。给定 $Y = y$ 的条件下, 条件 PDF $f_{X|Y}(x|y)$ 投影到 xz 平面上得到图 15。

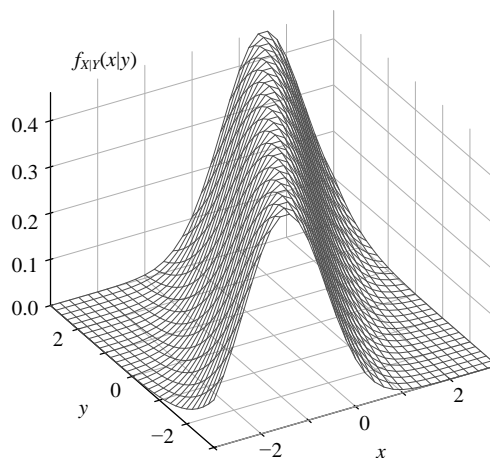
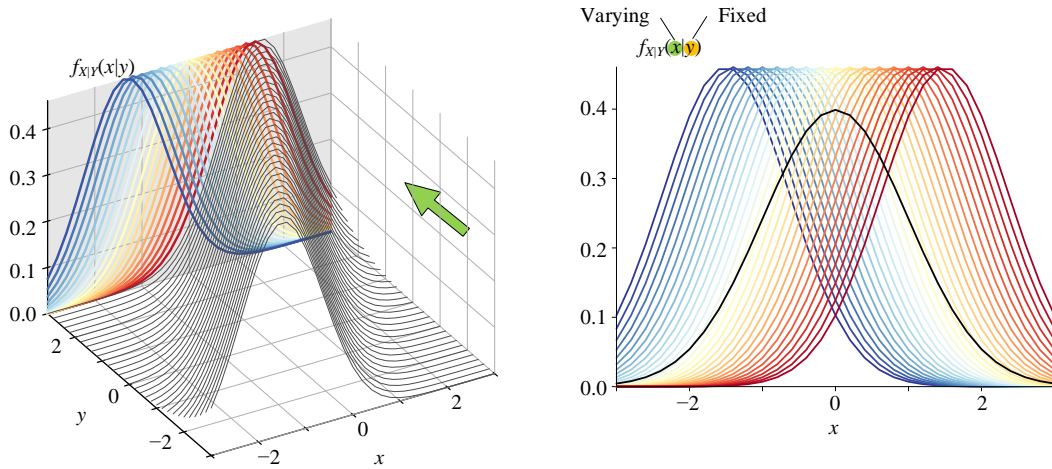
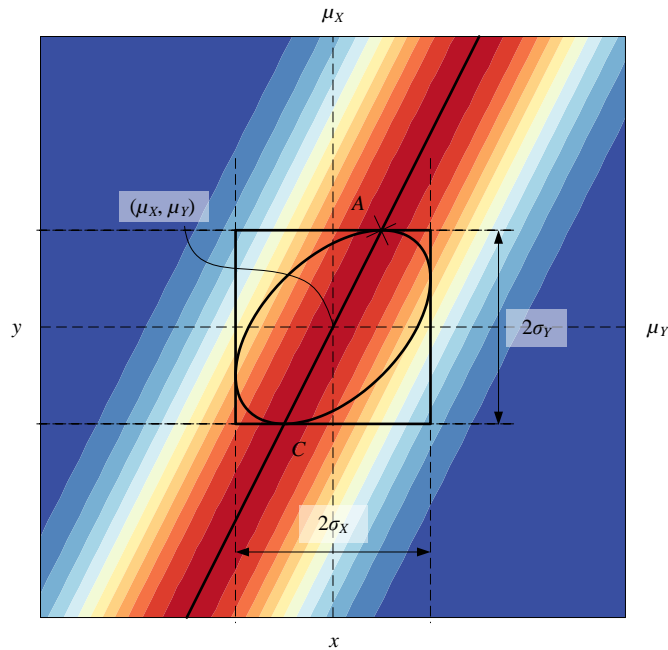


图 14. $f_{X|Y}(x|y)$ 曲面网格线图 15. $f_{X|Y}(x|y)$ 曲面在 xz 平面上投影

条件期望 $E(X|Y=y)$

图 16 所示为 $f_{X|Y}(x|y)$ 的平面等高线。图中的等高线都平行于给定 $Y=y$ 条件下， X 的条件期望 $E(X|Y=y)$ ，具体解析式为：

$$E(X|Y=y) = \mu_X + \rho_{X,Y} \frac{\sigma_X}{\sigma_Y} (y - \mu_Y) \quad (15)$$

图 16. $f_{X|Y}(x|y)$ 平面等高线

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

条件方差 $\text{var}(X|Y=y)$

给定 $Y=y$ 条件下， X 的条件**方差** $\text{var}(X|Y=y)$ 解析式为：

$$\text{var}(X|Y=y) = (1 - \rho_{X,Y}^2) \sigma_X^2 \quad (16)$$

给定 $Y=y$ 条件下， X 的条件**标准差** $\sigma_{X|Y=y}$ 解析式也是定值：

$$\text{std}(X|Y=y) = \sqrt{(1 - \rho_{X,Y}^2)} \cdot \sigma_X \quad (17)$$

图 17 所示为条件**标准差** $\sigma_{X|Y}$ 的几何含义。

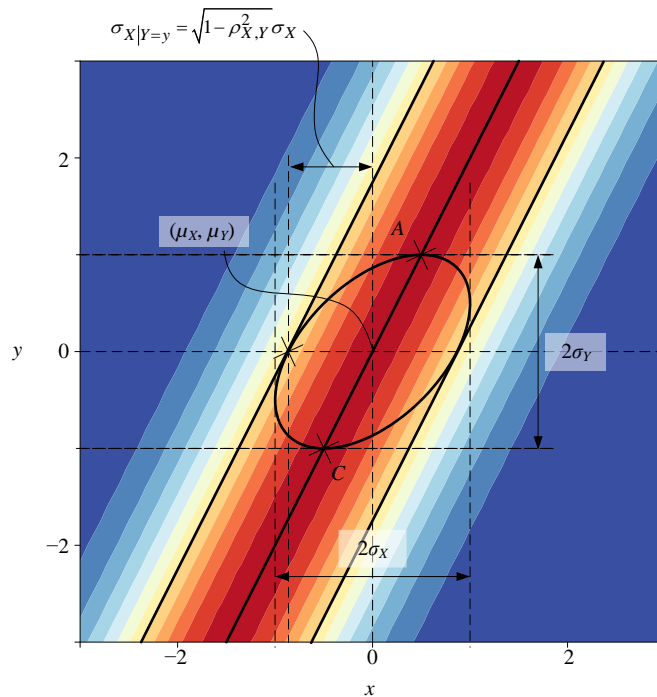


图 17. 条件**标准差** $\sigma_{X|Y}$ 的几何含义

以鸢尾花为例：条件期望 $E(X_1 | X_2 = x_2)$ 、**条件方差** $\text{var}(X_1 | X_2 = x_2)$

以鸢尾花花萼长度 (X_1)、花萼宽度 (X_2) 数据为例，假设 (X_1, X_2) 服从二元高斯分布。给定 $X_2 = x_2$ 条件下， X_1 的条件**期望** $E(X_1 | X_2 = x_2)$ 解析式为：

$$\begin{aligned}
 E(X_1 | X_2 = x_2) &= \mu_1 + \rho_{1,2} \frac{\sigma_1}{\sigma_2} (x_2 - \mu_2) \\
 &= 5.843 - 0.117 \times \frac{0.825}{0.434} (x_2 - 3.057) \\
 &= -0.222x_2 + 6.523
 \end{aligned} \tag{18}$$

条件**方差** $\text{var}(X_1 | X_2 = x_2)$ 解析式为：

$$\text{var}(X_1 | X_2 = x_2) = (1 - \rho_{1,2}^2) \sigma_1^2 \approx 0.671 \tag{19}$$

条件**标准差** $\sigma_{X_1 | X_2 = x_2}$ 解析式为定值：

$$\sigma_{X_1 | X_2 = x_2} = \sqrt{1 - \rho_{1,2}^2} \sigma_1 \approx 0.819 \tag{20}$$

类似地，如图 18 所示不管 x_2 怎么变，这个条件**标准差**为定值。

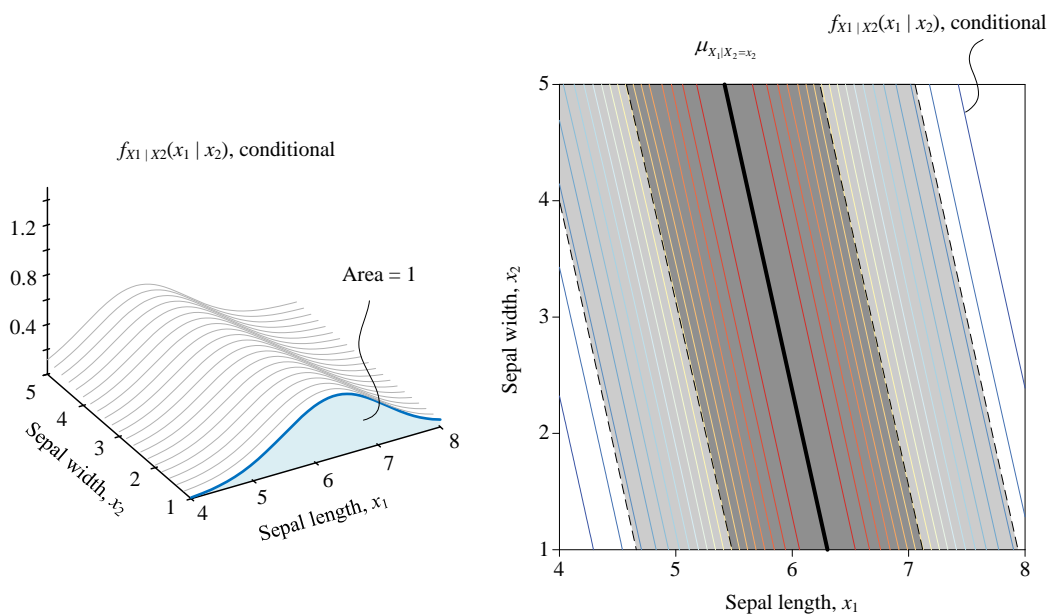
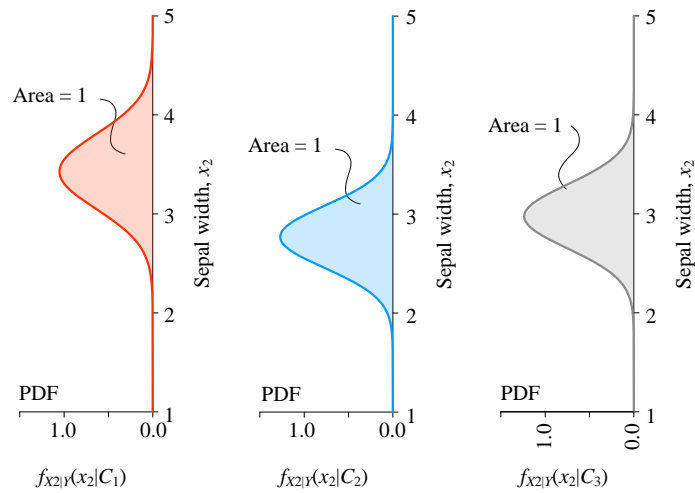


图 18. $f_{X_1 | X_2}(x_1 | x_2)$ 条件 PDF 密度三维等高线和平面等高线，不考虑分类

以鸢尾花为例，考虑标签

换个条件来看，如图 19 所示，给定鸢尾花分类条件，假设花萼宽度服从高斯分布。请大家自行计算给定鸢尾花分类为条件，花萼宽度的条件**期望** $E(X_2 | Y = C_k)$ 、条件**方差** $\text{var}(X_2 | Y = C_k)$ 。

图 19. 给定鸢尾花标签 Y ，花萼宽度的 PDF 曲线，连续随机变量

12.4 多元正态条件分布：引入矩阵运算

本节利用矩阵运算讨论多元正态条件分布。

多元高斯分布

如果随机变量向量 χ 和 γ 服从多维高斯分布：

$$\begin{bmatrix} \chi \\ \gamma \end{bmatrix} \sim N \left(\begin{bmatrix} \mu_\chi \\ \mu_\gamma \end{bmatrix}, \begin{bmatrix} \Sigma_{\chi\chi} & \Sigma_{\chi\gamma} \\ \Sigma_{\gamma\chi} & \Sigma_{\gamma\gamma} \end{bmatrix} \right) \quad (21)$$

其中， χ 为随机变量 X_i 构成的列向量， γ 为随机变量 Y_j 构成的列向量：

$$\chi = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_D \end{bmatrix}, \quad \gamma = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_M \end{bmatrix} \quad (22)$$

图 20 所示为多元高斯分布的均值向量、协方差矩阵形状。

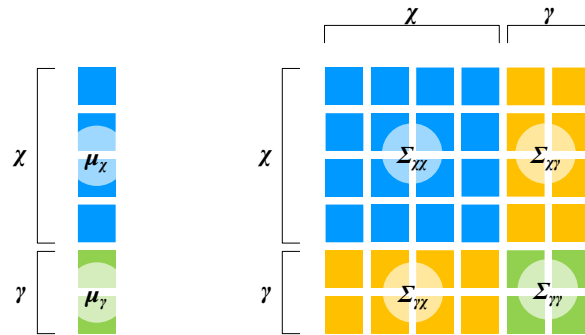


图 20. 均值向量、协方差矩阵形状

互协方差矩阵

注意， $\Sigma_{\gamma\chi}$ 的转置为 $\Sigma_{\chi\gamma}$ ：

$$(\Sigma_{\gamma\chi})^T = \Sigma_{\chi\gamma} \quad (23)$$

$\Sigma_{\chi\gamma}$ 也叫**互协方差矩阵** (cross-covariance matrix)，这是下一章要讨论的内容之一。

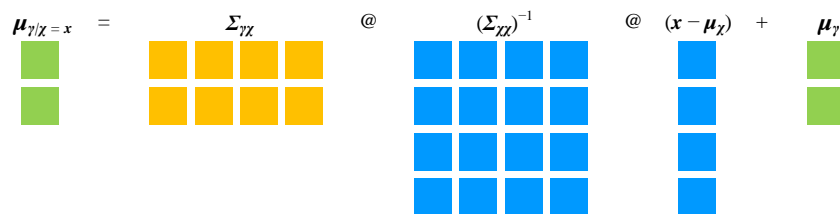
给定 $\chi = x$ 的条件

给定 $\chi = x$ 的条件下， γ 服从如下多维高斯分布：

$$\{\gamma | \chi = x\} \sim N \left(\underbrace{\Sigma_{\gamma\chi} \Sigma_{\chi\chi}^{-1} (x - \mu_\chi) + \mu_\gamma}_{\text{Expectation}}, \underbrace{\Sigma_{\gamma\gamma} - \Sigma_{\gamma\chi} \Sigma_{\chi\chi}^{-1} \Sigma_{\chi\gamma}}_{\text{Covariance matrix}} \right) \quad (24)$$

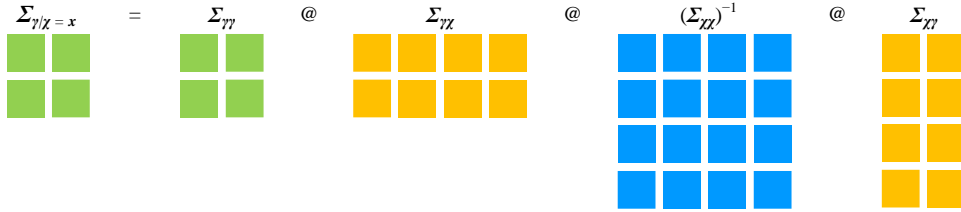
也就是说，如图 21 所示，给定 $\chi = x$ 的条件下 γ 的条件**期望**为：

$$E(\gamma | \chi = x) = \mu_{\gamma|\chi=x} = \Sigma_{\gamma\chi} \Sigma_{\chi\chi}^{-1} (x - \mu_\chi) + \mu_\gamma \quad (25)$$

图 21. 给定 $\chi = x$ 的条件下 γ 的期望值的矩阵运算

如图 22 所示，给定 $\chi = x$ 的条件下 γ 的**方差**为：

$$\Sigma_{\gamma|\chi=x} = \Sigma_{\gamma\gamma} - \Sigma_{\gamma\chi} \Sigma_{\chi\chi}^{-1} \Sigma_{\chi\gamma} \quad (26)$$

图 22. 给定 $\chi=x$ 的条件下 γ 的方差的矩阵运算

给定 $\gamma=y$ 的条件

同理，给定 $\gamma=y$ 的条件下 χ 服从如下多维高斯分布：

$$\{\chi|\gamma=y\} \sim N\left(\underbrace{\Sigma_{\chi\gamma} \Sigma_{\gamma\gamma}^{-1} (y - \mu_\gamma) + \mu_\chi}_{\text{Expectation}}, \underbrace{\Sigma_{\chi\chi} - \Sigma_{\chi\gamma} \Sigma_{\gamma\gamma}^{-1} \Sigma_{\gamma\chi}}_{\text{Covariance matrix}}\right) \quad (27)$$

即给定 $\gamma=y$ 的条件下 χ 的期望值为：

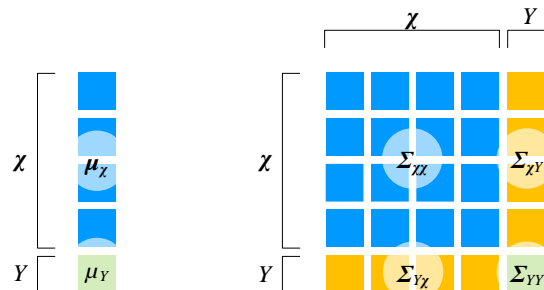
$$\mu_{\chi|\gamma=y} = \Sigma_{\chi\gamma} \Sigma_{\gamma\gamma}^{-1} (y - \mu_\gamma) + \mu_\chi \quad (28)$$

给定 $\gamma=y$ 的条件下 χ 的方差为：

$$\Sigma_{\chi|\gamma=y} = \Sigma_{\chi\chi} - \Sigma_{\chi\gamma} \Sigma_{\gamma\gamma}^{-1} \Sigma_{\gamma\chi} \quad (29)$$

单一因变量

特别地， γ 只有一个随机变量 Y 时，这对应线性回归中有多个自变量，只有一个因变量，如图 23 所示。

图 23. 均值向量、协方差矩阵形状， γ 只有一个随机变量

这种情况下，给定 $\chi = x$ 条件下 Y 的条件期望为：

$$\mu_{Y|\chi=x} = \Sigma_{Y\chi} \Sigma_{\chi\chi}^{-1} (x - \mu_{\chi}) + \mu_Y \quad (30)$$

(30) 对应多元线性回归。图 24 对应矩阵运算示意图。

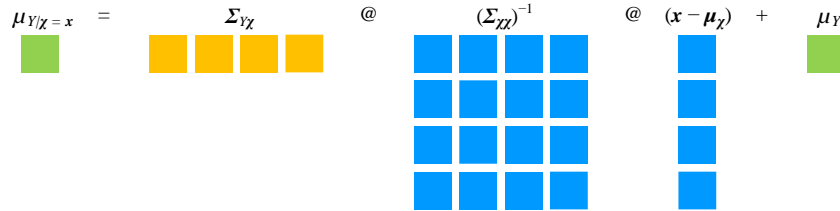


图 24. 给定 $\chi = x$ 条件下 Y 的条件期望

多元线性回归

不考虑常数项系数，如果是行向量表达的话，多元线性回归的系数 b 为：

$$b = [b_1 \quad b_2 \quad \cdots \quad b_D] = \Sigma_{Y\chi} \Sigma_{\chi\chi}^{-1} \quad (31)$$

图 25 所示为 b 的矩阵运算。

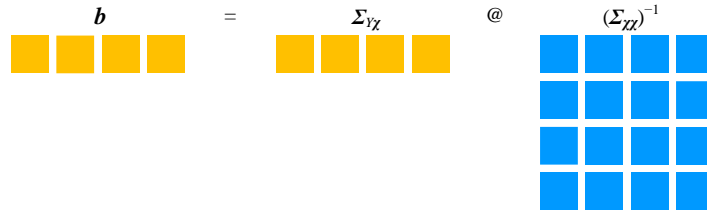


图 25. 计算多元回归的系数 b

常数项 b_0 为：

$$b_0 = -\Sigma_{Y\chi} \Sigma_{\chi\chi}^{-1} \mu_{\chi} + \mu_Y \quad (32)$$

简单线性回归

更特殊地，当 χ 和 Y 都只有一个随机变量时，即单一自变量 X 、单一因变量 Y ：

$$\mu_{Y|X=x} = \text{cov}(X, Y) (\sigma_X^2)^{-1} (x - \mu_X) + \mu_Y = \rho_{X,Y} \frac{\sigma_Y}{\sigma_X} (x - \mu_X) + \mu_Y \quad (33)$$

这和本书之前的 (8) 完全一致。本书第 24 章将接续讨论这一话题。

以鸢尾花为例

图 26 所示为鸢尾花数据的质心向量和协方差矩阵热图。我们用花萼长度、花萼宽度、花瓣长度为多元线性回归的多变量，用花瓣宽度为因变量。图 26 所示向量和协方差矩阵也据此分块。

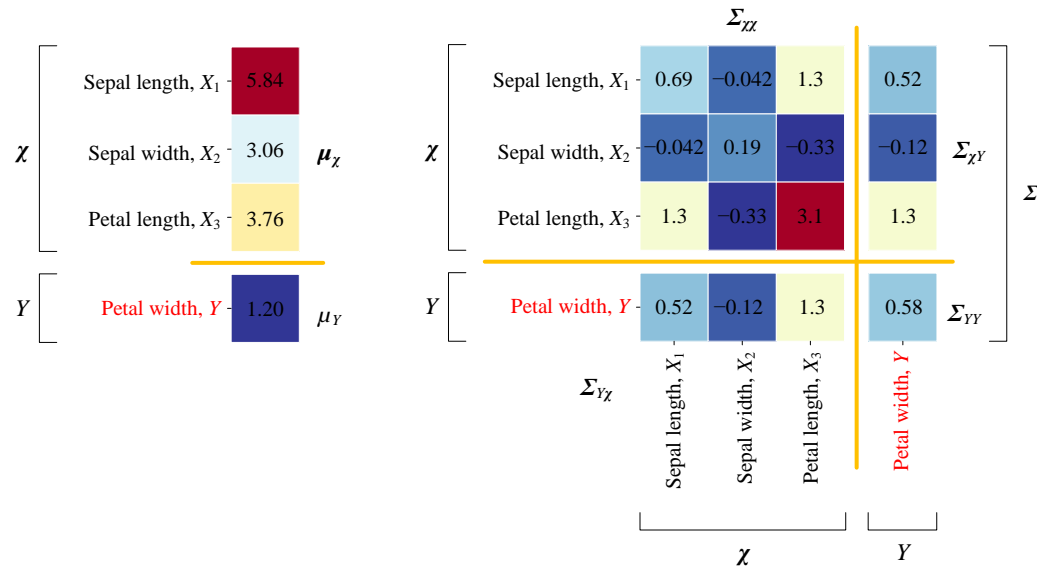


图 26. 质心向量、协方差矩阵热图

利用 (31)，我们可以计算得到回归系数：

$$\begin{aligned}
 \mathbf{b} &= [b_1 \quad b_2 \quad b_3] = \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_d \end{bmatrix} = \Sigma_{YX} \Sigma_{XX}^{-1} \\
 &= [0.516 \quad -0.122 \quad 1.296] \begin{bmatrix} 0.686 & -0.042 & 1.274 \\ -0.042 & 0.190 & -0.330 \\ 1.274 & -0.330 & 3.116 \end{bmatrix}^{-1} \\
 &= [-0.207 \quad 0.223 \quad 0.524]
 \end{aligned} \tag{34}$$

图 27 所示为上式运算的热图。

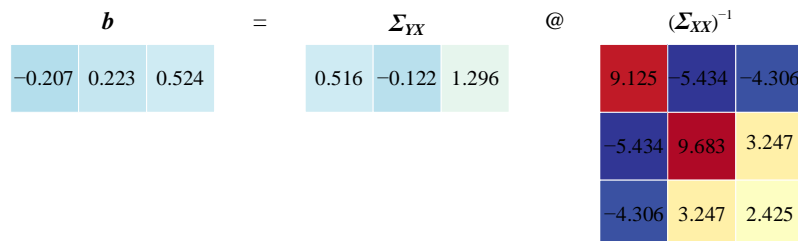


图 27. 矩阵计算系数向量 \mathbf{b}

利用 (32)，计算得到多元线性回归的常数项：

$$b_0 = 1.199 - [-0.207 \quad 0.223 \quad 0.524] \begin{bmatrix} 5.843 \\ 3.057 \\ 3.758 \end{bmatrix} = -0.24 \quad (35)$$

图 28 所示为对应计算。

$$b_0 = \mu_Y - \mathbf{b} @ \mu_X$$

b_0	=	μ_Y	-	\mathbf{b}	@	μ_X
-0.240		1.199		-0.207	0.223	0.524
					5.843	3.057
					3.758	

图 28. 矩阵计算常数 b_0

从而得到多元线性回归的解析式：

$$y = [0.516 \quad -0.122 \quad 1.296] \begin{bmatrix} 0.686 & -0.042 & 1.274 \\ -0.042 & 0.190 & -0.330 \\ 1.274 & -0.330 & 3.116 \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} - \begin{bmatrix} 5.843 \\ 3.057 \\ 3.758 \end{bmatrix} + 1.199 \quad (36)$$

$$= -0.207x_1 + 0.223x_2 + 0.524x_3 - 0.240$$

这个式子相当于用花萼长度、花萼宽度、花瓣长度作为变量估算花萼宽度。

有必要强调一下，线性回归并不一定表示因果关系。尽管线性回归可以用来探索变量之间的关系，但它并不会告诉我们一个变量是否是另一个变量的原因。因为在统计学中，相关性并不等于因果关系。

要确定两个变量之间的因果关系，需要进行实验研究，例如随机对照实验。在这种类型的实验中，研究人员可以控制潜在的影响因素，然后观察自变量对因变量的影响。因此，线性回归可以用于探索变量之间的关系，但要确定因果关系需要进行更深入的研究和分析。



Bk5_Ch12_03.py 绘制本节图像。



简单来说，条件高斯分布是指在已知某些变量的取值情况下，对另外一些变量的概率分布进行建模的一种方法。条件高斯分布在模式识别、机器学习、贝叶斯推断等领域都有广泛的应用。本节中大家看到条件高斯分布给线性回归提供一种全新的解读视角。