工程概率统计

Probability and Statistics for Engineering

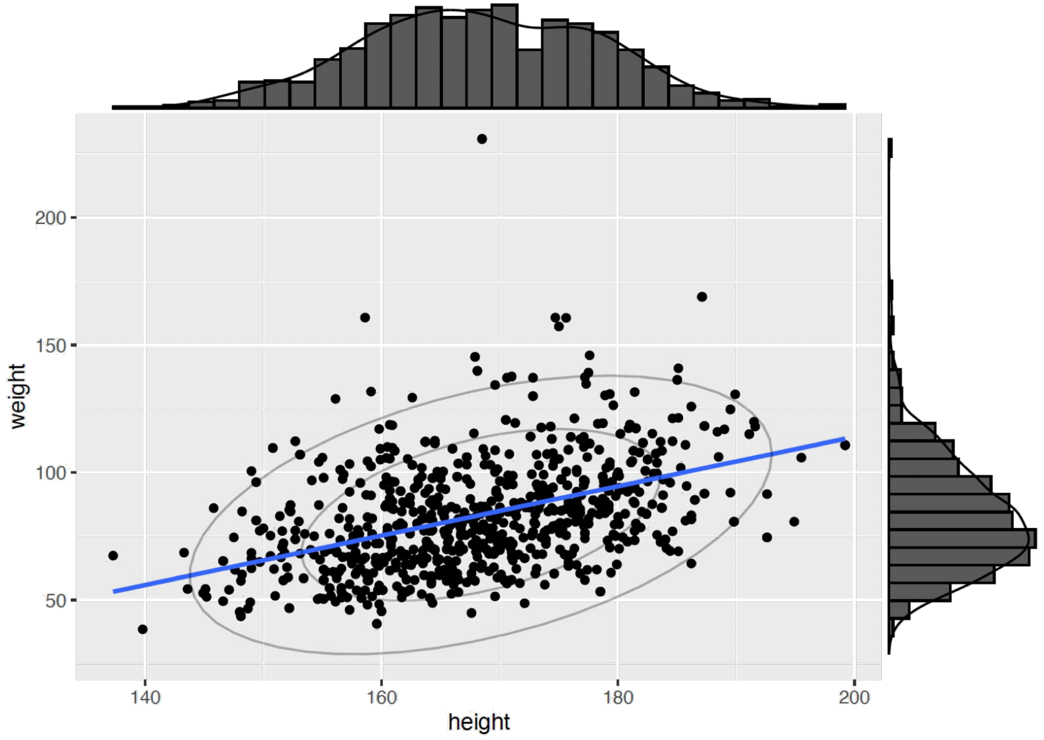
第三章 联合分布

Chapter 3 Joint Distributions

Chapter 3 Joint Distributions

3.4 Multivariate Normal Distribution

- In this section, we will talk about the bivariate normal distribution (二元正态分布), and the results can be generalized to multivariate normal distribution (多元正态分布).
- The bivariate normal/Gaussian distribution is commonly used to model the joint distribution of two normal random variables, particularly when they have some degree of linear relationship.
- Real-world examples:
  - Height and weight of adults
  - Father and son's heights
  - Test scores in two courses



3.4 Multivariate Normal Distribution

The Bivariate Normal Distribution

- Random vector (X, Y) is said to be bivariate normally distributed with means μ_X, μ_Y and variances σ_X^2 and σ_Y^2, and with correlation coefficient ρ, if the joint PDF of (X, Y) is given by $(-\infty < x, y < \infty)$

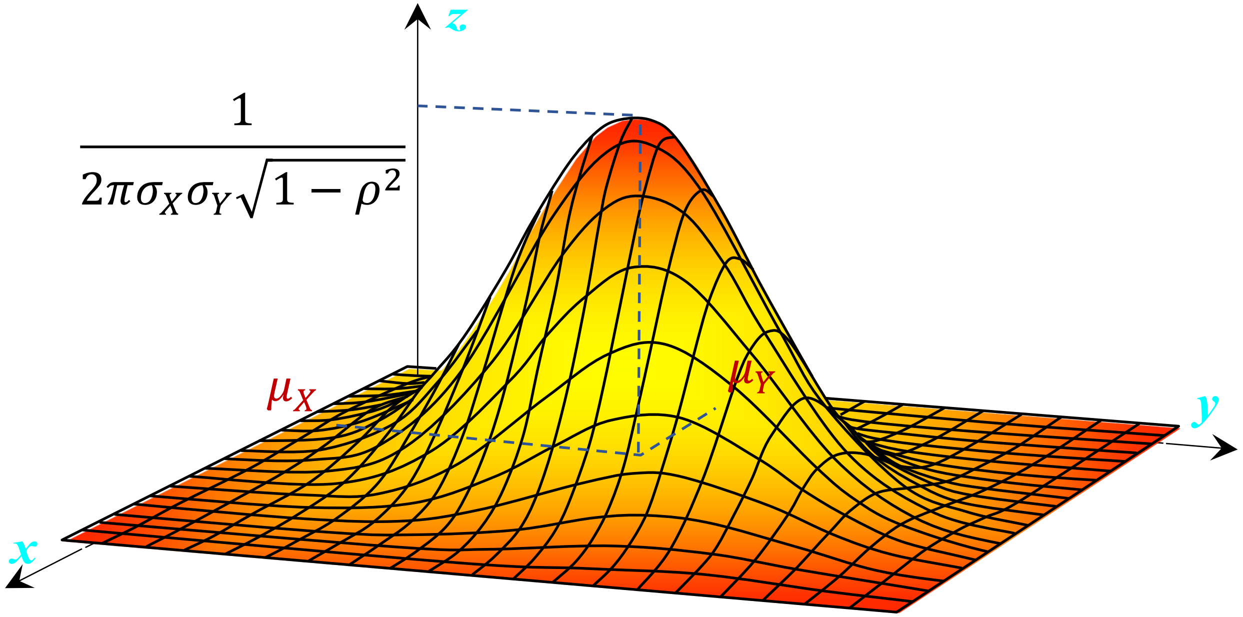
$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right).$$

- It can be expressed as

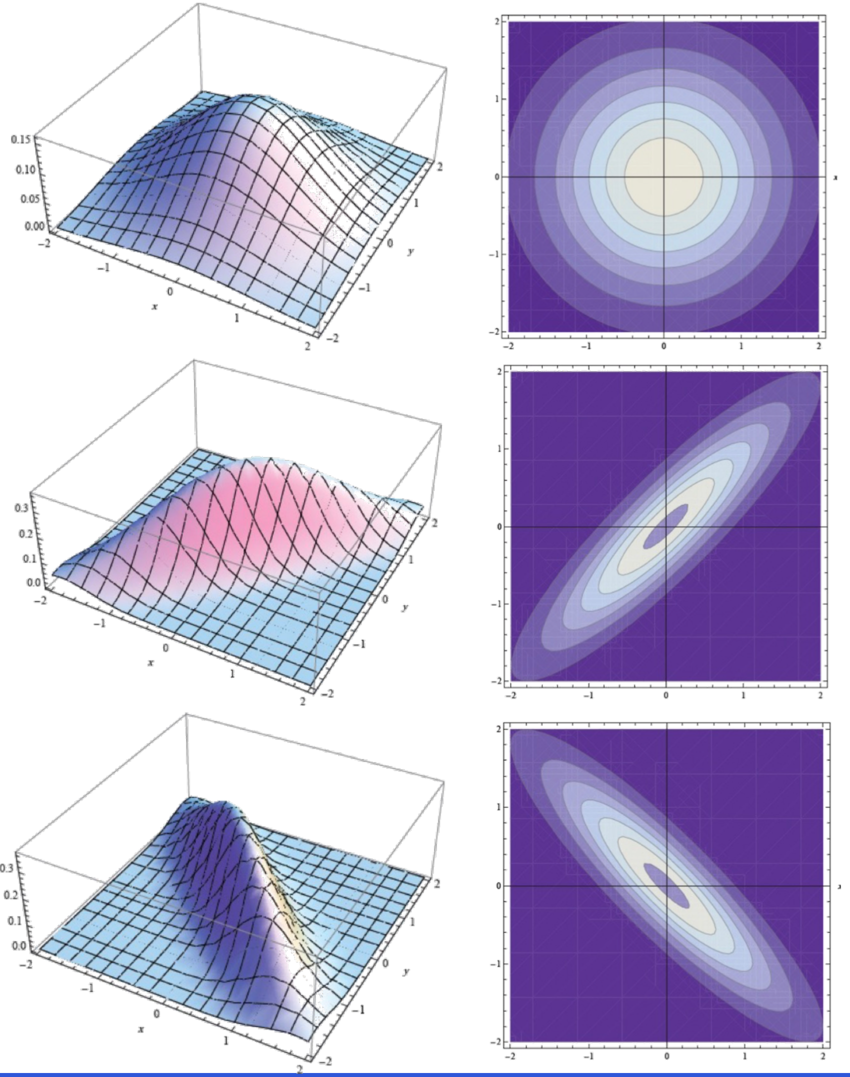
$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right) = \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

- where $\boldsymbol{\mu}$ is the mean vector (均值向量) and $\boldsymbol{\Sigma}$ is the variance-covariance matrix (方差-协方差矩阵).
- Specifically, if $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 1$, and $\rho = 0$, then it is said to be a standard bivariate normal distribution, i.e., $\mathrm{N}(\mathbf{0}, \mathbf{I})$.

3.4 Multivariate Normal Distribution



$\mu_X = \mu_Y = 0, \sigma_X = \sigma_Y = 1$



$\rho = 0$

$\rho = 0.9$

$\rho = -0.9$

3.4 Multivariate Normal Distribution

Marginal distributions are normal distributions.

- It can be shown that if $\binom{X}{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ($\boldsymbol{\mu} = \binom{\mu_X}{\mu_Y}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$), then $X \sim \mathrm{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$, and $\rho_{XY} = \rho$.

Proof:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{1}{\sqrt{1-\rho^2}}\right)^2 \left(\frac{y-\mu_Y}{\sigma_Y} - \frac{x-\mu_X}{\sigma_X}\right)^2} dy$$

$$\text{Let } t = \frac{1}{\left(\sqrt{1-\rho^2}\right)}\left(\frac{y-\mu_Y}{\sigma_Y} - \frac{x-\mu_X}{\sigma_X}\right) \Rightarrow dy = \sigma_Y\sqrt{1-\rho^2}\,dt$$

$$f_X(x) = \frac{1}{2\pi\sigma_X} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt = \frac{1}{2\pi\sigma_X} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} \cdot \sqrt{2\pi} = \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}}$$

$\therefore X \sim N(\mu_X, \sigma_X^2)$, and $Y \sim N(\mu_Y, \sigma_Y^2)$ follows similarly.

By variable substitution, details omitted.

$$\mathrm{Cov}(X,Y) = \mathrm{E}[(X-\mu_X)(Y-\mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y) f(x,y)dxdy = \rho\sigma_X\sigma_Y.$$

$$\Rightarrow \rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{\rho\sigma_X\sigma_Y}{\sigma_X\sigma_Y} = \rho.$$

3.4 Multivariate Normal Distribution

- Generally, if random variables X and Y are uncorrelated, then we not necessarily have X and Y are independent.

X and Y are independent $\Rightarrow$ X and Y are uncorrelated

X and Y are independent $\not\Leftarrow$ X and Y are uncorrelated

- However, uncorrelated does imply independent if X and Y jointly follow a bivariate normal distribution.

Proof: If (X, Y) follow a bivariate normal distribution and they are uncorrelated, i.e., $\rho = \rho_{XY} = 0$, then the joint PDF is

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right).$$

$$= \frac{1}{2\pi\sigma_X\sigma_Y} \exp\left(-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right)$$

$$= \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} \times \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}} = f_X(x) f_Y(y). \qquad \therefore X \text{ and } Y \text{ are independent.}$$

3.4 Multivariate Normal Distribution

Marginal distributions cannot uniquely determine the joint distribution!

Example 3.17

- Let r.v. $X \sim \mathrm{N}(0,1)$ and Z be a r.v. independent of X with PMF $\mathrm{P}(Z = 1) = \mathrm{P}(Z = -1) = 0.5$.
- Define $Y = ZX$, (1) show that $Y \sim \mathrm{N}(0,1)$; (2) show that X and Y are uncorrelated.

Solution

3.4 Multivariate Normal Distribution

Conditional distributions are normal distributions.

- Moreover, if $\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ $\left(\boldsymbol{\mu} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right)$, then

$$X|Y = y \sim \mathrm{N}\left(\mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y - \mu_Y), (1-\rho^2)\sigma_X^2\right), \quad Y|X = x \sim \mathrm{N}\left(\mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X), (1-\rho^2)\sigma_Y^2\right).$$

Proof:
$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$

$$= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right) \Big/ \frac{1}{\sqrt{2\pi}\sigma_Y} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_X} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{\rho^2(y-\mu_Y)^2}{\sigma_Y^2}\right]\right)$$

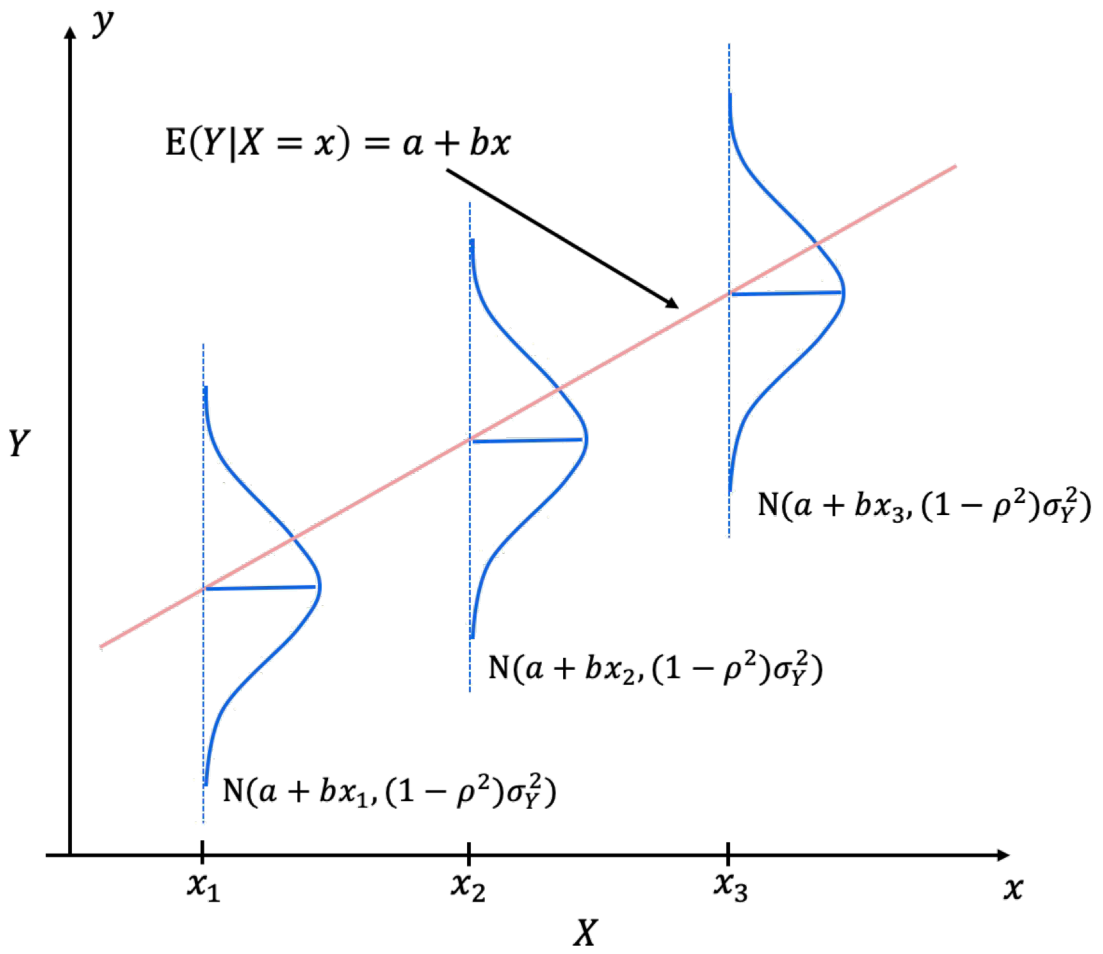
$$= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_X} \exp\left(-\frac{1}{2(1-\rho^2)\sigma_X^2}\left[(x-\mu_X) - \frac{\rho\sigma_X(y-\mu_Y)}{\sigma_Y}\right]^2\right)$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}\sigma_X} \exp\left(-\frac{1}{2(1-\rho^2)\sigma_X^2}\left[x - \left(\mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y-\mu_Y)\right)\right]^2\right)$$

$$\mathrm{E}(Y|X = x) = a + bx$$
$$b = \rho\frac{\sigma_Y}{\sigma_X}, a = \mu_Y - b\mu_X$$

3.4 Multivariate Normal Distribution

- Graphical illustration of the conditional distribution of Y given $X = x$.



- Y follows a normal distribution given any value of X.
- The mean of the normal distribution is a linear function of the value of X.
- These normal distributions have different means but the same variance.

3.4 Multivariate Normal Distribution

- Furthermore, if $\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ $\left(\boldsymbol{\mu} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right)$, then for any constants $c_1, c_2 \in \mathbb{R}$:

$$c_1X + c_2Y \sim \mathrm{N}(c_1\mu_X + c_2\mu_Y, c_1^2\sigma_X^2 + 2c_1c_2\rho\sigma_X\sigma_Y + c_2^2\sigma_Y^2)$$

Linear combinations still follow normal distributions.

Note: On Page 45, we provide the conclusion that a linear combination of independent normal random variables still follow a normal distribution.

The statement above provide a more general conclusion which does not require independence between the normal random variables, but require that their joint distribution is a bivariate (or. multivariate) normal distribution.

Proof of the statement is omitted here.

- More generally, for any real matrix $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, we have:

$$\boldsymbol{A}\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} a_{11}X + a_{12}Y \\ a_{21}X + a_{22}Y \end{pmatrix} \sim \mathrm{N}(\boldsymbol{A\mu}, \boldsymbol{A\Sigma A}^\top).$$

3.4 Multivariate Normal Distribution

Example 3.18

- Let X and Y denote the math score and the verbal score on the ACT college entrance exam of a randomly selected student.
- Previous history suggest that X and Y are bivariate normally distributed with means $\mu_X = \mu_Y = 22.7$, variances $\sigma_X^2 = 17.64$ and $\sigma_Y^2 = 12.25$, and correlation coefficient $\rho = 0.78$.
- Calculate:
- (1) The probability that a randomly selected student's math score is greater than 25?
- (2) The probability that a randomly selected student's math score is greater than 25 given that his/her verbal score is 25?
- (3) The probability that a randomly selected student has combined math and verbal score greater than 50?
- (4) The probability that a randomly selected student's math score is higher than his/her verbal score given that he/she has combined math and verbal score 50.

3.4 Multivariate Normal Distribution

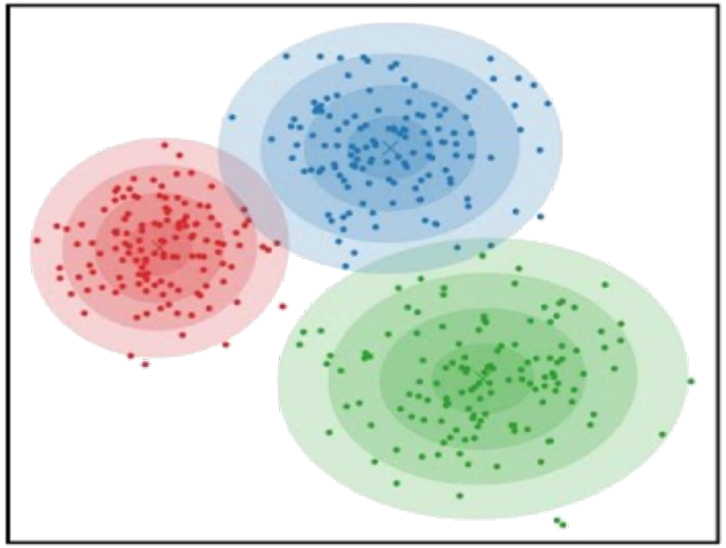
Solution

3.4 Multivariate Normal Distribution

Solution

3.4 Multivariate Normal Distribution

- Finally, we talk about an application of the multivariate normal distribution in machine learning, the **Gaussian Mixture Model (GMM, 高斯混合模型)**.
- GMM is a machine learning method used to determine the probability each data point belongs to a given cluster. It is a clustering method (聚类方法) used in unsupervised learning (非监督学习).
- Under GMM, the dataset is modeled as a mixture of several multivariate Gaussian distributions, assuming that individuals of different clusters come from different Gaussian distributions.
  - For a randomly selected individual, let Y be its cluster ID, which takes value in $\{1, 2, \ldots, K\}$, $\boldsymbol{X}$ be its feature vector, then $\boldsymbol{X}|Y = k \sim \mathrm{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$.
  - The goal is to infer $\mathrm{P}(Y = k|\boldsymbol{X} = \boldsymbol{x})$.



- The multivariate normal distribution has wide applications in pattern recognition, computer vision, natural language processing, signal processing, finance, and economics, etc.

End of Chapter 3