言語処理のための機械学習入門
1.4 連続確率変数

新納浩幸

連続確率変数

X の取り得る値は連続的な実数値

$$P(X = x) = 0$$

密度関数を利用して確率と関係づける

$$P(a < X < b) = \int_a^b \underline{f(x)}\,dx$$

$$\Rightarrow \quad \int_{-\infty}^{+\infty} f(x)\,dx = 1$$

平均と分散

$$E(X) = m = \int xf(x)dx$$

$$V(X) = \sigma^2 = \int (x-m)^2 f(x)dx$$

$$\Rightarrow \quad V(X) = E(X^2) - E(X)^2$$

分散公式

正規分布

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$

$$E(X) = \mu$$

$$V(X) = \sigma^2$$

$$X \sim N(\mu, \sigma^2)$$

多次元正規分布

$$f(\mathbf{x}) = \left(\frac{1}{2\pi}\right)^{p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$$

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_p)$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}$$

分散共分散行列

ディリクレ分布

n 次元確率変数の分布

$$\mathbf{x} = (x_1, x_2, \cdots, x_n)$$

$$0 \le x_i \le 1 \qquad x_1 + x_2 + \cdots + x_n = 1$$

$$f(\mathbf{x}; \boldsymbol{\alpha}) = \frac{1}{\int \prod_i x_i^{\alpha_i - 1} dx} \prod_i x_i^{\alpha_i - 1}$$