

Pima2021

Long Nguyen

July 26, 2021

Dữ liệu được đưa dưới dạng một danh sách các vector D chiều được ký hiệu là: $X = (\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n)^T$ với $\vec{x}_i \in \mathbb{R}^d$. Một phân phối chuẩn nhiều chiều định nghĩa bởi vector trung bình và $\vec{\mu}$ ma trận covariance Σ . Vector ngẫu nhiên \vec{X} được gọi là tuân theo phân phối đều D chiều ký hiệu là: $\vec{X} \sim N_D(\vec{\mu}, \Sigma)$, khi đó hàm mật độ xác suất có thể được tính như sau

$$f(\vec{x}; \vec{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp -\frac{1}{2}(\vec{x} - \vec{\mu})\Sigma^{-1}(\vec{x} - \vec{\mu})^T$$