Note for Probability Theory and Statistics

Linlin Ge

March 21, 2020

1 Distribution

Multivariate normal distribution

$$f_{\boldsymbol{x}}(x_1, \cdots, x_k) = \frac{e^{-\frac{(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}{2}}}{\sqrt{(2\pi)^k |\boldsymbol{C}|}}$$

Where,

$\boldsymbol{C}$ The symmetric covariance matrix.

$\boldsymbol{x}$ A real k-dimensional column vector.

2 Distance

2.1 Distance of Random Variables (Same Distribution, Covariance Matrix=$\boldsymbol{C}$)

Mahalanobis Distance

(1) Distance of vector $\boldsymbol{x}$ and set:

$$D_M(\boldsymbol{x}) = \sqrt{(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

If $\boldsymbol{x}$ is a k-dimensional Gaussian random vector with mean vector $\boldsymbol{\mu}$ and rank k covariance matrix $\boldsymbol{C}$, then the Mahalanobis distance $D_M^2(\boldsymbol{x})$ follows a χ^2-distribution with d degrees of freedom.

(2) Distance of two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

3 Error Function

3.1 Approximation

$$erf(x) \approx 1 - \frac{1}{(1 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4)^4} \quad x \geq 0$$

Where $a_1 = 0.278393, a_2 = 0.230389, a_3 = 0.000972, a_4 = 0.078108$, and its maximum error: 5×10^{-4}

References