Normal distribution

Univariate. A random variable ξ is said to have a normal distribution with mean μ and variance σ^2 if its probability density function (pdf) is given by

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

where μ is the mean and σ^2 is the variance of the distribution. More compactly, it can be written as

$$\xi \sim \mathcal{N}(\mu, \sigma^2) \tag{2}$$

Uncorrelated multivariate. A random vector $\boldsymbol{\xi} = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_k \end{pmatrix}$ is said to have an uncorrelated multivariate normal distribution with mean $\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix}$ and variances $\sigma_1^2, \ldots, \sigma_k^2$ if the pdf of every random component of $\boldsymbol{\xi}$ is given by

$$f_{\xi_j}(x) = \frac{1}{\sigma_j\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2\right\} \tag{3}$$

where μ_j is the mean and σ_j^2 is the variance of the j-th component of the.

All components of $\boldsymbol{\xi}$ are assumed to be independent, so the joint pdf of $\boldsymbol{\xi}$ is the product of the pdfs of its components:

$$\begin{aligned} f_{\boldsymbol{\xi}}(x_1, \ldots, x_k) &= \prod_{i=1}^{k} f_{\xi_i}(x_i) \\ &= \prod_{i=1}^{k} \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right\} \end{aligned} \tag{4}$$

Covariance matrix. All variance parameters $\sigma_1^2, \ldots, \sigma_k^2$ can be combined into a covariance matrix Σ. The covariance matrix is a symmetric positive definite matrix that describes the covariance between the components of $\boldsymbol{\xi}$.

$$\Sigma = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_k^2 \end{pmatrix} \tag{5}$$

Here, the covariance matrix is diagonal (all off-diagonal elements are zero), because we assumed that the components of $\boldsymbol{\xi}$ are uncorrelated, *i.e.*, $\mathrm{Cov}[\xi_i, \xi_j] = 0$ for all $i \neq j$.

The pdf of the multivariate normal distribution can be written in terms of the covariance

$$f_{\boldsymbol{\xi}}(x_1, \ldots, x_k) = \frac{\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}}{\sqrt{(2\pi)^k \det \Sigma}} \tag{6}$$

The covariance matrix Σ above is a diagonal matrix, but in general, it's a symmetric positive definite matrix that describes the covariance between the components of $\boldsymbol{\xi}$:

$$\Sigma := \begin{pmatrix} \mathrm{Cov}[\xi_1, \xi_1] & \ldots & \mathrm{Cov}[\xi_1, \xi_k] \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}[\xi_k, \xi_1] & \ldots & \mathrm{Cov}[\xi_k, \xi_k] \end{pmatrix}. \tag{7}$$

If we substitute the non-diagonal covariance matrix Σ into the pdf, we get the general form of the multivariate normal distribution.

Technically, each component of Σ is the covariance between the corresponding components

$$\Sigma_{i,j} := \mathrm{Cov}[\xi_i, \xi_j] = \mathbb{E}\left[(\xi_i - \mu_i)(\xi_j - \mu_j)\right]. \tag{8}$$

For a sample $X = \{x_1, \ldots, x_\ell\} \subset \mathbb{R}$, the variance is the average of the squared differences from the mean:

$$\mathrm{Var}[X] := \frac{1}{\ell} \sum^{\ell} (x_i - \bar{x})^2$$

The term $\det \Sigma$ is the generalized variance.

Mahalanobis distance. The distance between a point $\boldsymbol{x}$ and the distribution $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ can be measured using the Mahalanobis distance.

The premise is that the covariance matrix Σ captures the correlations between the components of $\boldsymbol{\xi}$. The Mahalanobis distance is a measure of how many standard deviations away a point $\boldsymbol{x}$ is from the mean $\boldsymbol{\mu}$, taking into account the correlations between the components of $\boldsymbol{\xi}$.

We can define a quadratic form

$$\begin{aligned} Q(\boldsymbol{x}) &:= (\boldsymbol{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \\ &= \sum_{i,j} (x_i - \mu_i) \big(\mathrm{Cov}[\xi_i, \xi_j]\big)^{-1} (x_j - \mu_j). \end{aligned} \tag{9}$$

it can be interpreted

Quadratic form $Q(\boldsymbol{x})$ is a scalar function of a vector $\boldsymbol{x}$ that can be expressed as as weighted sum of the squares of the components of $\boldsymbol{x}$:

$$Q(\boldsymbol{x}) = \sum_{i,j} w_{i,j} x_i x_j.$$

These weights can be gathered into a matrix W, and the quadratic form can be written as a matrix product:

$$Q(\boldsymbol{x}) = \boldsymbol{x}^\top W \boldsymbol{x}.$$