Multivariate Normal Distribution

In this lesson we discuss the multivariate normal distribution. We begin with a brief reminder of basic concepts in probability for random variables that are scalars and then generalize them for random variables that are vectors.

Basic concepts in Probability

Let $x \in R$ be a random variable. The expectation of x is $E(x) = \sum_x x \Pr(x)$ for discrete variables and $E(x) = \int xp(x)dx$ for continuous variables (where $p(x)$ is the probability density function of x). The expectation is a linear operator: let $y \in R$ be another random variable, then $E(ax + y) = aE(x) + E(y)$.

The variance of x is $\text{var}(x) = E((x - E(x))^2)$. Notice that $\text{var}(x) \geq 0$ and that $\text{var}(x) = E(x^2) - E(x)^2$. The standard deviation of x is $\sigma = \sqrt{\text{var}(x)}$.

Let $y \in R$ be another random variable. The covariance of x and y is $cov(x, y) = E((x - E(x))(y - E(y)))$ (notice that $\text{cov}(x, x) = \text{var}(x)$). Notice that $\text{var}(x + y) = \text{var}(x) + \text{var}(y) + 2\text{cov}(x, y)$ and that $\text{cov}(x, y) = E(xy) - E(x)E(y)$. If $cov(x, y) = 0$ we say that x and y are uncorrelated. If x and y are independent then x,y are uncorrelated. The opposite is not true in general.

Let $x \in R$. $x \sim N(\mu, \sigma^2)$ (x is normally distributed with parameters μ and σ^2) if

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

where $p(x)$ is the probability density function of x. μ is the expectation of x and σ^2 is the variance of x.

Multivariate Normal Distribution

Generalization for vector random variables: definitions

Let $x = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \in R^N$. The expectation of x is $E(x) = \begin{bmatrix} E(x_1) \\ \vdots \\ E(x_N) \end{bmatrix} \in R^N$.
The covariance matrix of x is $Cov(x) = E((x-\mu)(x-\mu)^T) \in R^{N\times N}$ where $\mu = E(x) \in R^N$. In other words, the entries of the covariance matrix are $Cov(x)_{i,j} = Cov(x_i, x_j)$. Notice that the covariance matrix is symmetric and positive definite and $Cov(x) = E(xx^T) - E(x)E(x)^T$

Let $x \in R^N$. $x \sim N(\mu, \Sigma)$ (x is normally distributed with parameters μ and Σ) if

$$p(x) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} e^{-\frac{(x-\mu)^T\Sigma^{-1}(x-\mu)}{2}} \tag{2}$$

$\mu \in R^N$ is the mean and $\Sigma \in R^{N\times N}$ is symmetric and positive definite. Σ is the covariance matrix of x, i.e. $\Sigma_{i,j} = Cov(x_i, x_j)$.

Understanding the multivariate normal distribution

Let $x_1, x_2 \in R$ be two independent random variables $x_1 \sim N(\mu_1, \sigma_1^2), x_2 \sim N(\mu_2, \sigma_2^2)$. The joint distribution of x_1, x_2 is:

$$p(x_1, x_2) = p(x_1)p(x_2) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} = \tag{3}$$

$$\frac{1}{(2\pi)^{2/2}(\sigma_1^2\sigma_2^2)^{1/2}} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_2)^2}{2\sigma_2^2}} = \frac{1}{(2\pi)^{2/2}(\sigma_1^2\sigma_2^2)^{1/2}} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_2)^2}{2\sigma_2^2}}$$

Let $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$. Σ is a diagonal matrix, thus $\Sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix}$. Since x_1, x_2 are independent, $\Sigma = Cov(x)$.

$(x-\mu)^T\Sigma^{-1}(x-\mu) = (\frac{x_1-\mu_1}{\sigma_1^2}, \frac{x_2-\mu_2}{\sigma_2^2}) \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix} = \frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}$. With these notations we can write the joint distribution $p(x_1, x_2)$ as:

$$\frac{1}{(2\pi)^{2/2}(\sigma_1^2\sigma_2^2)^{1/2}} e^{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_2)^2}{2\sigma_2^2}} = \frac{1}{(2\pi)^{2/2}\det(\Sigma)^{1/2}} e^{-\frac{(x-\mu)^T\Sigma^{-1}(x-\mu)}{2}} \tag{4}$$

which shows that the joint distribution is a 2D normal distribution with parameters μ and Σ.

The equiprobable curves (curves along which the probability density is constant) of the joint distribution are ellipses: the only term of $p(x_1, x_2)$ that depends on x_1, x_2 is $\frac{(x_1-\mu_1)^2}{2\sigma_1^2} + \frac{(x-\mu_2)^2}{2\sigma_2^2}$. Setting this positive term equal to a constant, we get the equation of an ellipse with center at (μ_1, μ_2) and horizontal and vertical axes. The ratio between the length axes of the ellipse is σ_1/σ_2.

Suppose that we apply a rotation on x_1, x_2, we expect to see the ellipse of equal probability rotated correspondingly. Let U be a 2D rotation matrix. Applying a rotation on x_1, x_2 introduces dependencies: Suppose that σ_1 is very large and σ_2 is very small, then x is distributed close to (and along) the horizontal axis. For example, if U is a rotation of $\pi/4$, then Ux is approximately distributed close to the line $y = x$ (which means that the two coordinates of Ux are strongly dependent).

We will now calculate the distribution of $y = Ux$ using the formula:

$$y = f(x) \Rightarrow p(y) = p(x)\left|\frac{dx}{dy}\right| = p(f^{-1}(y))\left|\frac{df^{-1}(y)}{dy}\right| \tag{5}$$

for one-to-one mapping f.
The determinant of a rotation matrix is 1 so we need only substitute $x = U^T y$ in the exponent:

$$\begin{aligned} p(y) = p(x(y)) = \frac{1}{(2\pi)^{2/2}|\Sigma|^{1/2}} e^{-\dfrac{(U^T y-\mu)^T\Sigma^{-1}(U^T y-\mu)}{2}} = \\ \frac{1}{(2\pi)^{2/2}1\cdot|\Sigma|^{1/2}\cdot 1} e^{-\dfrac{(y-U\mu)^T U\Sigma^{-1}U^T(y-U\mu)}{2}} = \\ \frac{1}{(2\pi)^{2/2}|U|^{1/2}\cdot|\Sigma|^{1/2}\cdot|U^T|^{1/2}} e^{-\dfrac{(y-\mu_y)^T((U^T)^{-1}\Sigma U^{-1})^{-1}(y-U\mu_y)}{2}} = \\ \frac{1}{(2\pi)^{2/2}|U\Sigma U^T|^{1/2}} e^{-\dfrac{(y-\mu_y)^T(U\Sigma U^T)^{-1}(y-U\mu_y)}{2}} = \\ \frac{1}{(2\pi)^{2/2}|\Sigma_y|^{1/2}} e^{-\dfrac{(y-\mu_y)^T\Sigma_y^{-1}(y-U\mu_y)}{2}} \end{aligned} \tag{6}$$

where $\mu_y = U\mu$ is the expectation of y and $\Sigma_y = U\Sigma U^T$ is the covariance matrix of y. A similar computation shows that for a general regular A, $y = Ax$ is distributed normally with $\mu_y = Ax$ and $\Sigma_y = A\Sigma A^T$ (for a general A, its determinant should be considered) .

Can we always present a normally distributed vector as a linear transformation of independent scalar normal random variables ? In particular, can we represent it as an orthonormal distribution of such variables ? Can we always view the normal distribution as an n-dimensional ellipse ?

The answer to all these question is yes. To see this we will use the Singular Value Decomposition of the covariance matrix Σ. Since Σ is symmetric and positive definite, its SVD is $\Sigma = UDU^T$ where U is an $n \times n$ orthonormal matrix and D is an $n \times n$ diagonal matrix whose diagonal entries positive. This means that any normally distributed vector $y \sim N(\mu_y, \Sigma_y)$ can be written as $y = Ux$ where $x \sim N(U^T\mu_y, D)$ and the coordinates of x, x_i are independent and normally distributed with variance d_i (the i^{th} entry on the diagonal of D).