0.1 Perspective of Probability

0.1.1 Abstract

In this issue, we will observe multivariate Gaussian distribution from the perspective of probability.

0.1.2 Prior Knowledge

$$
\begin{aligned}
x &\backsim N(\mu, \sigma^2) \\
\mu &\in \mathcal{R}^p, \sigma \in \mathcal{R}^p \\
x_i &\backsim N(\mu_i, \sigma_i) \\
p(x_i) &= \frac{1}{\sqrt{2\pi}\sigma_i} \exp(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2})
\end{aligned}
$$

0.1.3 Derivation

First, let's assume that each x_i is *iid*(*independent identically distribution*) as below:

$$
\begin{aligned}
p(x) &= \prod_{i=1}^{p} p(x_i) \\
&= \frac{1}{(2\pi)^{\frac{p}{2}} \prod_{i=1}^{p} \sigma_i} \exp(-\frac{1}{2} \sum_{i=1}^{p} (\frac{(x_i-\mu_i)^2}{\sigma_i^2})) \\
&= \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2} \begin{pmatrix} x_1-\mu_1 & x_2-\mu_2 & ... & x_p-\mu_p \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 & ... & 0 \\ ... & ... & ... & ... \\ 0 & ... & 0 & \frac{1}{\sigma_p^2} \end{pmatrix} \begin{pmatrix} x_1-\mu_1 \\ ... \\ ... \\ x_p-\mu_p \end{pmatrix}) \\
&= \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu))
\end{aligned}
$$

The above is the probability density function of multivariate Gaussian distribution.

We know that σ is a positive semidefinite matrix, so we can perform singular value decomposition. So we have:

$$\begin{aligned}
\Sigma &= UVU^T \\
&= \begin{pmatrix} u_1 & ... & u_p \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & ... & 0 \\ ... & 0 & ... & ... \\ 0 & ... & ... & \lambda_p \end{pmatrix} \begin{pmatrix} u_1^T \\ . \\ . \\ u_p^T \end{pmatrix} \\
&= \begin{pmatrix} u_1\lambda_1 & ... & u_p\lambda_p \end{pmatrix} \begin{pmatrix} u_1^T \\ . \\ . \\ u_p^T \end{pmatrix} \\
&= \sum_{i=1}^{p} u_i \lambda_i u_i^T
\end{aligned}$$

then

$$\begin{aligned}
\Sigma^{-1} &= (UVU^T)^{-1} \\
&= (U^T)^{-1} V^{-1} U^{-1} \\
&= UV^{-1}U^T \\
&= \sum_{i=1}^{p} u_i \frac{1}{\lambda_i} u_i^T
\end{aligned}$$

Let's set $\Delta = (x-\mu)^T \Sigma^{-1} (x-\mu)$

Substitute the results derived above into:

$$\begin{aligned}
\Delta &= (x-\mu)^T \Sigma^{-1} (x-\mu) \\
&= (x-\mu)^T \sum_{i=1}^{p} u_i \frac{1}{\lambda_i} u_i^T (x-\mu) \\
&= \sum_{i=1}^{p} (x-\mu)^T u_i \frac{1}{\lambda_i} u_i^T (x-\mu)
\end{aligned}$$

Let's set $y_i = (x-\mu)^T u_i$

Here, y_i represents the coordinate value of x projected onto the new orthogonal basis u_i after centralization.
so:

$$\Delta = \sum_{i=1}^{p} \frac{y_i^2}{\lambda_i}$$

Next, let's look at the probability density function of multivariate Gaussian distribution:

$$p(x) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu))$$

You can see that only the exponential part of the formula is related to the variable x. The previous factor is to make the probability sum 1.
Therefore, the probability of Gaussian distribution is directly related to the value of Δ.

We assume $p = 2$, then:

$$\frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2} = \Delta$$

We were surprised to find that this is very similar to the elliptic equation. The value of Δ is not fixed, so for different x, these sample points form concentric ellipses in the plane. This is one of the properties of Gaussian distribution.