MLE for 1D Gaussian PDF

假設我們有一組純量 $x_i, i=1\cdots n$，如果這些點的分佈會趨近一個中間值，同時左右接近對稱，則我們可用一維高斯密度函數 $g\left(x_i;\mu,\sigma^2\right)$ 來描述產生這些點的機率密度函數（PDF, probability density function）：

$$g\left(x;\mu,\sigma^2\right)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

其中 μ 代表此密度函數的中心點，σ^2 則約略代表此密度函數的寬窄，這些參數決定了此密度函數的特性。

我們的目標，是要求得 PDF 的最佳參數 $\left[\mu,\sigma^2\right]$ 以描述所觀察到的資料點。在上述一維高斯密度函數的假設下，當 $x=x_i$ 時，其機率密度為 $g\left(x_i;\mu,\sigma^2\right)$，若我們假設 $x_i, i=1\sim n$ 之間為互相獨立的事件，則發生 $X=\{x_1,x_2\cdots x_n\}$ 的機率密度為

$$p\left(X;\mu,\sigma^2\right)=\prod_{i=1}^{n}g\left(x_i;\mu,\sigma^2\right)$$

由於 X 是已經發生之事件，因此我們希望找出 $\left[\mu,\sigma^2\right]$ 值，使得 $p\left(X;\mu,\sigma^2\right)$ 能有最大值，此種估測 PDF 參數的方法，即稱為**最佳似然率估測法**（MLE, maximum likelihood estimation）。（注意：$p\left(X;\mu,\sigma^2\right)$ 不是機率，而是機率密度，有時候又稱為 likelihood 或似然率。）

欲求得 $p\left(X;\mu,\sigma^2\right)$ 的最大值，我們通常將之轉化為求下列 $J(\mu,\sigma^2)$ 的最大值

$$\begin{aligned}
J(\mu,\sigma^2)&=\ln p\left(X;\mu,\sigma^2\right)\\
&=\ln\left[\prod_{i=1}^{n}g\left(x_i;\mu,\sigma^2\right)\right]\\
&=\sum_{i=1}^{n}\ln g\left(x_i;\mu,\sigma^2\right)\\
&=\sum_{i=1}^{n}\left[-\frac{1}{2}\ln\left(2\pi\right)-\ln\sigma-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]\\
&=-\frac{n}{2}\ln\left(2\pi\right)-n\ln\sigma-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2
\end{aligned}$$

欲求最佳的 μ 及 σ 值，直接求 $J(\mu,\sigma^2)$ 對 μ 及 σ 的微分，並令其為零：

$$\frac{\partial J(\mu,\sigma^2)}{\partial \mu} = \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right) = 0 \Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\frac{\partial J(\mu,\sigma^2)}{\partial \sigma} = -\frac{n}{\sigma} - \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)\left(-\frac{x_i-\mu}{\sigma^2}\right) = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i-\hat{\mu}\right)^2$$

上述所得到的 $\hat{\mu}$ 及 $\hat{\sigma}^2$ 值，即是使用 MLE 原則所得到的最佳參數值。