最大似然估计

Guangyao Zhao

2022-02-25

Contents

似然函数

似然性（likelihood）和概率（possibility）同样可以表示事件发生的可能性大小，但是两者有很大区别：

- 概率：是在已知参数 θ 的情况下，发生观测结果 x 可能性大小。
- 似然性：从观测结果 x 出发，分布函数的参数 θ 的可能性大小。

似然函数如下：

$$\mathrm{L}(\theta|x) = p(x|\theta)$$

其中：x 已知，θ 未知。若对于两个参数 θ_1 和 θ_2 ，有：

$$\mathrm{L}(\theta_1|x) = p(x|\theta_1) > p(x|\theta_2) = \mathrm{L}(\theta_2|x)$$

那么意味着 $\theta = \theta_1$ 的时候，随机变量 X 生成 x 的概率大于当参数 $\theta = \theta_2$ 。这也正是似然函数的意义所在。若观测数据为 x ，那么 θ_1 比 θ_2 更有可能是分布函数的参数。

最大似然估计

最大似然函数的思想在于，对于给定观测数据 x ，希望能反推出所有参数 $\theta_1, \theta_2, ..., \theta_k$ 中找出能最大概率生成观测数据的参数 θ^* 作为估计结果。即，被估计的参数 θ^* 应该满足：

$$\mathrm{L}(\theta^*|x) = p(x|\theta^*) > p(x|\theta) = \mathrm{L}(\theta|x), \theta = \theta_1, ..., \theta_k$$

在实际运算中，将待估参数 θ 作为变量，计算生成观测数据 x 的概率函数 $p(x|\theta)$ ，并通过求导找到最大概率函数的参数即可：

$$\theta^* = \underset{\theta}{\operatorname{argmax}}\, p(x|\theta)$$

离散型随机变量的最大似然估计

在参数 θ 下，分布函数随机取到 $x_1, x_2, ..., x_n$ 的概率为：

$$P(X|\theta) = \prod_{i=1}^{n} p(x_i; \theta)$$

构造似然函数：

$$\mathrm{L}(\theta|X) = p(X|\theta) = \prod_{i=1}^{n} p(x_i; \theta)$$

似然函数是一个关于 θ 的函数，要找到最大概率生成 x 的参数，即需要找到 $\mathrm{L}(\theta|x)$ 取最大值时的 θ。此时需要对其求导：

$$\frac{d}{d\theta}\mathrm{L}(\theta|x) = 0$$

因为很一般情况下式子是累积形式 (独立同分布)，所以可借助对数函数简化问题：

$$\frac{d}{d\theta}\ln(\mathrm{L}(\theta|x)) = 0$$

上式通常称作对数似然方程。如果包含多个参数 $\theta_1, \theta_2, ..., \theta_k$，则可对多个参数分别求导联立方程组。

高斯分布下的最大似然函数

假设样本符合高斯分布，即 $X \sim N(\mu, \sigma^2)$，其中 μ 为均值，σ 为方差。$x_1, x_2, ...x_n$ 为来自 X 的一组观察值，求 μ 和 σ 的最大似然估计。

X 的概率密度函数：

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp - \frac{(x-\mu)^2}{2\sigma^2}$$

观察值 $x_1, x_2, ...x_n$ 的似然函数为：

$$\mathrm{L}(\theta|X) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp - \frac{(x_i-\mu)^2}{2\sigma^2}$$

$$\ln \mathrm{L}(\theta|X) = -n \ln \sqrt{2\pi}\sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

对其求导得：

$$\frac{d}{d\mu} \ln \mathrm{L}(\theta|X) = \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0$$

$$\frac{d}{d\sigma} \ln \mathrm{L}(\theta|X) = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i - \mu)^2 = 0$$

最后求解得：

$$\mu = \bar{x}$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$