对高斯分布的贝叶斯推断

emacsun

目录

1 简介

在最大似然的框架中，我们得到了高斯分布μ和$\mathbf{\Sigma}$的点估计。现在，我们采用贝叶斯方法处理这个问题，为此我们需要引入这些参数的先验估计。

2 假设方差已知

作为一个例子，我们考虑单个高斯随机变量x，假设方差σ^2已知，我们要做的是从N个观测值$\mathbf{X} = \{x_1, \ldots, x_N\}$中推断出均值$\mu$。似然函数定义为：

$$p(\mathbf{X}|\mu) = \prod_{n=1}^{N} p(x_n|\mu) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{n=1}^{N} (x_n - \mu)^2 \right\} \tag{2.1}$$

值得强调的是：似然函数$p(\mathbf{X}|\mu)$不是关于μ的概率密度函数，并且这个似然函数不是归一化的。

从式(2.1)我们可以看出似然函数中μ的二次型出现在指数位置上。如果我们选择高斯分布作为μ的先验分布$p(\mu)$。那么这个高斯分布就是这个似然函数的共轭先验分布，因为对应的后验分布是两个关于μ的指数二次型函数的乘积。因

此，我们假设μ的先验分布为：

$$p(\mu) = \mathcal{N}(\mu|\mu_0, \sigma_0^2) \tag{2.2}$$

因此后验分布为：

$$p(\mu|\mathbf{X}) \propto p(\mathbf{X}|\mu)p(\mu) \tag{2.3}$$

通过化简我们得到：

$$p(\mu|\mathbf{X}) = \mathcal{N}(\mu|\mu_N, \sigma_N^2) \tag{2.4}$$

其中：

$$\mu_N = \frac{\sigma^2}{N\sigma_0^2 + \sigma^2}\mu_0 + \frac{N\sigma_0^2}{N\sigma_0^2 + \sigma^2}\mu_{ML} \tag{2.5}$$

$$\frac{1}{\sigma_N^2} = \frac{1}{\sigma_0^2} + \frac{N}{\sigma^2} \tag{2.6}$$

$$\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n \tag{2.7}$$

我们对式 (2.5)中的后验均值和方差做简单的分析。首先，我们注意到μ_N是μ_0和μ_{ML}的一个折中。如果观测到的点数$N = 0$，那么μ_N就是μ_0。当$N \to \infty$时，μ_N趋向于μ_{ML}。同样的，对于σ_N^2，我们发现使用倒数的表达更清晰易懂。我们称σ_N^2为方差，称$\frac{1}{\sigma_N^2}$为精度。精度具有可加性，并且后验精度是先验精度加上每一次观测数据的精度。当我们增加观测点数的时候，精度逐渐增加，对应的方差逐渐降低。如果没有观测数据，我们得到的就是先验精度，如果$N \to \infty$，方差$\sigma_N^2 \to 0$，此时后验分布会在最大似然解处形成一个无穷高的尖峰。因此，我们通过贝叶斯估计，我们得到了μ点估计的最大似然解。注意对于有限的N，如果我们假设$\sigma_0^2 \to \infty$,那么后验估计μ_N同样收敛到最大似然解，此时后验方差$\sigma_N^2 = \sigma^2/N$

我们之前 看到了过高斯分布的均值可以采用迭代的方式计算得出。实际上，贝叶斯估计也可以采用这种方式。

$$p(\mu|D) \propto \left[p(\mu)\prod_{n=1}^{N} p(\mathbf{x}_n|\mu)\right]p(\mathbf{x}_N|\mu) \tag{2.8}$$

式 (2.8)中在中括号中的项代表观测到$N-1$个数据之后的后验分布。我们看到，这个观测值可以当做第N次观测的先验分布。实际上，贝叶斯推断的这种序贯视角可以用在任何观测数据是独立同分布的场景中。

3　假设方差未知

截止目前，我们的处理手段是方差已知，然后对μ做估计。那么，如果方差未知怎么办呢？假设均值已知，我们来估计方差。和本文之前一样，我们也提出一个关于$\lambda = 1/\sigma^2$的先验估计。关于λ的似然函数可以写为：

$$p(\mathbf{X}|\lambda) = \prod_{n=1}^{N} \mathcal{N}(x_n|\mu, \lambda^{-1}) \propto \lambda^{N/2} \exp\left\{ -\frac{\lambda}{2} \sum_{n=1}^{N} (x_n - \mu)^2 \right\} \tag{3.1}$$

对应的相应的共轭先验应该具有的形式：λ的指数幂乘以λ的线性函数。这样的描述与gamma分布非常类似。gamma分布的定义为：

$$\mathrm{Gam}(\lambda|a, b) = \frac{1}{\Gamma(a)} b^a \lambda^{a-1} \exp(-b\lambda) \tag{3.2}$$

这里$\Gamma(a)$是gamma函数，出现在这里是为了保证gamma分布是归一化的。gamma分布的期望和方差为：

$$\mathbb{E}[\lambda] = \frac{a}{b} \tag{3.3}$$

$$\mathrm{var}[\lambda] = \frac{a}{b^2} \tag{3.4}$$

此处我们考虑先验分布$\mathbf{Gam}(\lambda|a_0, b_0)$，然后我们得到后验概率：

$$p(\lambda|\mathbf{X}) \propto \lambda^{a_0 - 1} \lambda^{N/2} \exp\left\{ -b_0\lambda - \frac{\lambda}{2} \sum_{n=1}^{N} (x_n - \mu)^2 \right\} \tag{3.5}$$

显然，式 (3.5)是一个新的gamma分布，参数为$\mathrm{Gam}(\lambda|a_N, b_N)$，其中：

$$a_N = a_0 + \frac{N}{2} \tag{3.6}$$

$$b_N = b_0 + \frac{1}{2} \sum_{n=1}^{N} (x_n - \mu)^2 = b_0 + \frac{N}{2} \sigma_{ML}^2 \tag{3.7}$$

其中σ_{ML}^2是方差的最大似然估计。

对于式 (3.6)，我们发现经过N次观测之后，a的值增加了$N/2$。因此我们可以把a_0视为$2a_0$次有效的先验观测。同样，对于式 (3.7)，我们发现经过N次观测，b的值增加了$\frac{N}{2}\sigma_{ML}^2$。所以我们可以把b_0解释为先验的$2a_0$次观测的等效方差$\frac{b_0}{a_0}$。

4　方差和均值都未知

当方差和均值都未知时，为了找到一个共轭先验分布，我们考虑μ和λ的先验函数：

$$p(\mathbf{X}|\mu, \lambda) = \prod_{n=1}^{N} \left(\frac{\lambda}{2\pi} \right)^{1/2} \exp\left\{ -\frac{\lambda}{2} (x_n - \mu)^2 \right\} \tag{4.1}$$

上式可以近似为：

$$\left[\lambda^{1/2}\exp(-\tfrac{\lambda\mu^2}{2})\right]^N \exp\left\{\lambda\mu\sum_{n=1}^{N}x_n-\frac{\lambda}{2}\sum_{n=1}^{N}x_n^2\right\} \tag{4.2}$$

我们要找到一个共轭先验分布$p(\mu,\lambda)$其形式和式 (4.2)类似，即：

$$p(\mu,\lambda) = \left[\lambda^{1/2}\exp(-\tfrac{\lambda\mu^2}{2})\right]^{\beta}\exp\{c\lambda\mu-d\lambda\} \tag{4.3}$$

$$= \exp\left\{-\frac{\beta\lambda}{2}(\mu-c/\beta)^2\right\}\lambda^{\beta/2}\exp\left\{-(d-\frac{c^2}{2\beta})\lambda\right\} \tag{4.4}$$

其中c,d,β是常量。根据$p(\mu,\lambda)=p(\mu|\lambda)p(\lambda)$，通过观察式 (4.3),我们令$p(\mu|\lambda)$是一个高斯分布，其精度是$\lambda$的线性函数;令$p(\lambda)$是一个gamma分布。所以：

$$p(\mu,\lambda)=\mathcal{N}(\mu|\mu_0,(\beta\lambda)^{-1})\mathrm{Gam}(\lambda|a,b) \tag{4.5}$$

其中，我们定义$\mu_0=c/\beta,a=1+\beta/2,b=d-c^2/2\beta$。式 (4.5)又被称为正态gamma分布或者高斯gamma分布。注意式(4.5)不是简单的正态分布和gamma分布的乘积。因为μ的精度是λ的线性函数。即使我们选择了一个先验分布保证μ和λ是独立的，在迭代过程中得到的后验分布中μ和λ也会纠缠在一起。