

分布推定・モデル選択

データ D を観測した時、それを生成した分布を推定したい。候補となる分布(確率密度関数)の集合 \mathbb{M} から、観測したデータを生成した確率が最も高いモデル \mathcal{M} を見つけたい、という問題を考える。

データ D が与えられた下でのモデル \mathcal{M} の事後確率 $p(\mathcal{M}|D)$ が最も高くなるようなモデル $\hat{\mathcal{M}}$ は、(1)式のように書ける。

$$\hat{\mathcal{M}} = \operatorname{argmax}_{\mathcal{M}} p(\mathcal{M}|D) \quad (1)$$

(1)式の事後確率 $p(\mathcal{M}|D)$ はベイズの定理から(2)式のように変形できる。さらに、データの確率 $p(D)$ は最早モデル \mathcal{M} に関係しないの定数であるから(3)式のようになる。

$$p(\mathcal{M}|D) = \frac{p(D|\mathcal{M})p(\mathcal{M})}{p(D)} \quad (2)$$

$$\propto p(D|\mathcal{M})p(\mathcal{M}) \quad (3)$$

ここで、モデルの事前分布 $p(\mathcal{M})$ を一様分布($p(\mathcal{M}_1) = p(\mathcal{M}_2) = \dots = p(\mathcal{M}_{|\mathbb{M}|}) = \text{const.}$)と仮定する。また、候補のモデルは少数のパラメータ θ (分布の平均とか分散など)で規定されるパラメトリックモデルと仮定する。すると、(5)式のようになり、これは $p(D|\theta, \mathcal{M})$ を θ で周辺化している形だから、周辺尤度(Marginal Likelihood, ML)と呼ぶ。このように、モデル \mathcal{M} の事後確率 $p(\mathcal{M}|D)$ は周辺尤度に比例するから、周辺尤度が最大となるモデルを見つければ良い。(対数をとった対数周辺尤度(log Marginal Likelihood, logML)の最大化や、対数周辺尤度にマイナスをつけたベイズ自由エネルギー(Bayesian Free Energy, BFE)の最小化を考えることもある。)

$$p(\mathcal{M}|D) \propto p(D|\mathcal{M})p(\mathcal{M}) \quad (3)$$

$$\propto p(D|\mathcal{M}) \quad (4)$$

$$= \int p(D|\theta, \mathcal{M})p(\theta|\mathcal{M})d\theta = \text{Marginal Likelihood} \quad (5)$$

$p(D|\theta, \mathcal{M})$ はあるモデル \mathcal{M} がパラメータ θ を取った時のデータ D の尤度、 $p(\theta|\mathcal{M})$ はあるモデル \mathcal{M} のもとでのパラメータ θ の事前分布である。尤度や事前分布をうまいこと設定すると周辺尤度は解析的に計算できる(うまいこと設定せずとも近似的に周辺尤度を計算する方法も存在する)。

例えば、モデルをガウス分布 $\mathcal{N}(\mu, \sigma)$ 、モデルのパラメータ μ の事前分布をガウス分布 $\mathcal{N}(\nu, \tau)$ とする(σ は既知と仮定)。ただし、 $D = \{x_i\}_{i=1}^N$ 、 $\theta = \mu$ 。仮定(6)(7)式より(5)式のインテグラルの中身を計算すると(8)式になる。(5)式は(8)式を μ に対して積分するから、積分されるのは $\mathcal{N}(\bar{\mu}, b)$ で、(9)式のようになる。(9)式には既知数しかないので計算できる。

$$x_i|\mu, \mathcal{M} \sim \mathcal{N}(\mu, \sigma) \quad (6)$$

$$\mu|\mathcal{M} \sim \mathcal{N}(\nu, \tau) \quad (7)$$

$$\begin{aligned} p(D|\theta, \mathcal{M})p(\theta|\mathcal{M}) &= \left[\prod_{i=1}^N p(x_i|\mu, \mathcal{M}) \right] p(\mu|\mathcal{M}) \\ &= \left[\prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \right] \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\mu - \nu)^2}{2\tau^2}\right) \\ &= \frac{1}{\sqrt{(2\pi)^N \sigma^{2(N-1)} (\sigma + N\tau^2)}} \exp\left[\frac{1}{2} \left(\frac{(\sigma^2\nu + N\bar{x}\tau^2)^2}{\sigma^2\tau^2(\sigma^2 + N\tau^2)} - \frac{\nu^2\sigma^2 + \bar{x}^2\tau^2}{\sigma^2\tau^2} \right) \right] \\ &\quad \sqrt{\frac{\sigma^2 + N\tau^2}{2\pi\sigma^2\tau^2}} \exp\left[-\frac{N\tau^2 + \sigma^2}{2\sigma^2\tau^2} \left(\mu - \frac{N\tau^2\bar{x} + \sigma^2\nu}{N\tau^2 + \sigma^2} \right)^2 \right] \\ &= \frac{1}{\sqrt{(2\pi)^N \sigma^{2(N-1)} (\sigma + N\tau^2)}} \exp\left[\frac{1}{2} \left(\frac{(\sigma^2\nu + N\bar{x}\tau^2)^2}{\sigma^2\tau^2(\sigma^2 + N\tau^2)} - \frac{\nu^2\sigma^2 + \bar{x}^2\tau^2}{\sigma^2\tau^2} \right) \right] \mathcal{N}(\bar{\mu}, b) \end{aligned} \quad (8)$$

$$\begin{aligned} \int p(D|\theta, \mathcal{M})p(\theta|\mathcal{M})d\theta &= \frac{1}{\sqrt{(2\pi)^N \sigma^{2(N-1)} (\sigma + N\tau^2)}} \exp\left[\frac{1}{2} \left(\frac{(\sigma^2\nu + N\bar{x}\tau^2)^2}{\sigma^2\tau^2(\sigma^2 + N\tau^2)} - \frac{\nu^2\sigma^2 + \bar{x}^2\tau^2}{\sigma^2\tau^2} \right) \right] \int \mathcal{N}(\bar{\mu}, b)d\mu \\ &= \frac{1}{\sqrt{(2\pi)^N \sigma^{2(N-1)} (\sigma + N\tau^2)}} \exp\left[\frac{1}{2} \left(\frac{(\sigma^2\nu + N\bar{x}\tau^2)^2}{\sigma^2\tau^2(\sigma^2 + N\tau^2)} - \frac{\nu^2\sigma^2 + \bar{x}^2\tau^2}{\sigma^2\tau^2} \right) \right] \end{aligned} \quad (9)$$