最大熵和最小鉴别原理

赵丰

2018 年 6 月 13 日

最大熵原理问题的提法：某离散随机变量 X，其概率分布 $p(x)$ 未知，已知若干函数在 $p(x)$ 下的期望：$\sum_{x\in\mathcal{X}} p(x)f_m(x) = C_m, m = 1, 2, \ldots, M$，求最佳估计 $\hat{p}(x)$。

求解原理是取概率分布的熵为目标函数 $H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x)$，而

$$\hat{p}(x) = \arg\max_{p(x)} H(X) \tag{1}$$

欠定约束下最大熵分布满足如下形式：

$$\hat{p}(x) = \exp\left[-\lambda_0 - \sum_{m=1}^{M} \lambda_m f_m(x)\right] \tag{2}$$

其中参数 $\lambda_0, \lambda_i (i = 1, \ldots, M)$ 由 $M+1$ 个约束条件确定（包含 $\sum_{x\in\mathcal{X}} p(x) = 1$）。

对于连续型随机变量，用积分代替求和，约束条件为

$\int_S p(x) = 1, \int_S p(x)f_m(x) = C_m, m = 1, 2, \ldots, M$ 则最大熵分布同 (2)式。

最小鉴别信息原理：欠定问题且有先验分布（$q(x)$）。

取概率分布的鉴别信息（相对熵）为目标函数：$D(p||q) = \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)}$

$$\hat{p}(x) = \arg\min_{p(x)} D(p||q) \tag{3}$$

用 Lagrange 乘子法可以求出欠定约束下最小鉴别信息分布为满足如下形式：

$$\hat{p}(x) = q(x)\exp\left[\lambda_0 + \sum_{m=1}^{M} \lambda_m f_m(x)\right] \tag{4}$$