

# 统计物理模型与算法介绍

## 1 最大熵估计

最大熵估计 MEE 是根据概率分布的熵 (Entropy) 最大化的原则来推断概率分布的形式。

Let  $\mathcal{C} = \{C_1, C_2, \dots, C_n, \dots\}$  代表 Collection(Ensemble);

Let  $m_i = \{m_i(C_1), m_i(C_2), \dots, m_i(C_n), \dots\}$  代表观测值,  $i = 1, 2, \dots, r$ 。

$P(C_i)$  代表  $C_i$  出现的概率, 其满足:

$$\sum_C P(C) = 1. \quad (1)$$

$m_i^*$  是已知的平均值。即有:

$$\sum_C P(C) \cdot m_i(C) = m_i^*. \quad (2)$$

$S = -\sum_C P(C) \log(P(C))$  为概率分布  $P(C)$  的熵。最大化熵 (用拉格朗日乘子法) 知  $P(C)$  的分布为:

$$P(C) = \frac{1}{Z} e^{H(C)}. \quad (3)$$

其中  $Z = \sum_C e^{H(C)}$ ,  $H(C) = \sum_{i=1}^r \theta_i m_i(C)$ 。  $\theta_i$  为参数, 通过已知的平均值  $m_i^*$  求出来。

**注** 以下的原则能用来得到最优函数:

1. 最大熵原理: 最优概率分布
2. 最短描述长度
3. 最小作业量原理: 最优函数

**注** 一下有一些原则能用来对生成式概率模型估计参数:

1. 误差最小化
2. 极大似然估计、EM 算法:  $\max p$
3. 变分推断 (最小化 KL 散度):  $KL(q||p)$ , 用  $q(x)$  逼近  $p(x)$
4. 能量最低原理 (Gibbs 自由能):  $E = -\log(p)$ ,  $p$  最大,  $E$  最小

## 2 Metropolis 准则: 以概率来接受新状态

假设当前系统的构型为  $C_n$ , 由于系统受到一定的干扰(随机性), 系统的构型会发生改变, 假设改变后的构型为  $C_{n+1}$ , 对应的, 系统能量由  $E_n$  变成  $E_{n+1}$ 。定义系统由  $C_n$  到  $C_{n+1}$  的接受概率为  $p$ :

$$p = \begin{cases} 1, & E_{n+1} < E_n; \\ \exp(-\beta(E_{n+1} - E_n)), & E_{n+1} \geq E_n. \end{cases} \quad (4)$$

其中,  $\beta = \frac{1}{T}$  表示反温度,  $T$  表示系统温度。

由第1节的玻尔兹曼分布可知: 构型的能量越低, 其存在的概率越大。所以, 如果构型改变后的能量变低了, 这种改变以概率 1 被接受; 否则, 将以构型改变后的存在概率  $\exp(-\beta E_{n+1})$  除以构型改变前的存在概率  $\exp(-\beta E_n)$  的可能性来接受这种改变。可以用电子跃迁来想象为何, 电子从高能级一定可以跃向低能级, 但从低能级跃迁到高能级的可能性就是其在高能级的概率除以低能级的概率, 这种跃迁的可能性就是我们定义的接受概率。

我们定义的接受概率实际上是在给定温度  $T$  下系统是如何受到既有确定性(能量最小化)又有随机性(干扰)的影响来发生改变的。这是一种总体上是确定性又带有随机性的策略。

## 3 模拟退火

我们依据 Metropolis 准则来叙述模拟退火算法:

1. 在给定温度  $T$  下, 随机改变系统的状态, 应用 Metropolis 准则以概率  $p$  来接受新状态;
2. 降低温度, 进行一次步骤 1; 直到温度达到给定温度为止。

### 例 3.1 硬币实验

1.  $\beta(\beta = 1/T)$ : 反温度;
2.  $\{\sigma_i\}(\sigma_i \in \{-1, 1\}, i = 1..N)$ : 硬币初始状态;
3.  $E = \sum_{i=1}^N \sigma_i$ : 硬币总能量;
4. 随机选择一个  $i$ , 尝试改变  $i$  硬币的状态  $\sigma_i$ :

$$\sigma_i \rightarrow -\sigma_i$$

$$E \rightarrow E'$$

$$\Delta E = E' - E$$

5. 以  $P_{flip} = \min\{1, e^{-\beta \Delta E}\}$  的概率翻转硬币  $i$ ;
6. 重复 4, 5 直到  $\beta$  增加到给定的  $\beta_{final}$ 。

## 4 马尔科夫链蒙特卡罗 MCMC

MCMC 是用来给定已知的概率分布,采一些样本是其服从给定的分布。

在此小节,我们将限制用 MCMC 采取由最大熵估计得到的概率分布。

**例 4.1** 随机图 MCMC 采样

1. 给定平均的边数  $\langle E \rangle$ , 解出参数  $\theta: \langle E \rangle = \binom{N}{2} \frac{e^\theta}{1+e^\theta}$ ;
2. 随机生成一个图  $G$ , 再随机选取一对节点, 改变边的状态 (右边变无边, 没变边右边), 得到新图  $G'$ ;
3. 用 Metropolis 准则来接受新图  $G'$ , 概率为  $p = \min\{1, e^{H(G')-H(G)}\}$ ; 若接受新图, 则选取其作为样本, 否则, 取原图作为样本;
4.  $G_1 \rightarrow G_2 \rightarrow G_3 \rightarrow \dots$