

2019-2020 第二学期 《计算统计》- 期中大作业

(2 班, 大作业, 共 45 分, 1.5 小时)

1. (共 30 分) 不完全数据 logistic 回归的 EM 算法: 假设

$$Y_i \sim \text{Bernoulli}(p_i)$$

$$p_i = \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}$$

其中, x_1, \dots, x_{50} 为已知常数, 变量 Y 's 的最后 10 个数字为缺失数据。

- 1). 区分已观测数据 X 和缺失数据 Z . (6 分)
- 2). 给出 EM 算法中牵涉到的未知参数 θ . (6 分)
- 3). 给出完整数据的对数似然函数. (6 分)
- 4). 基于当前 θ , 即 $\hat{\theta}^{(t)}$, 请写出基于已经观测变量和 $\hat{\theta}^{(t)}$ 的对数似然函数。

(12 分)

2. (共 15 分) 假设 Y_1, \dots, Y_n 为独立同分布的变量, 其概率密度函数为

$$f(\theta) = (\theta^2 + \theta)(1 - y)^{\theta-1}y, \quad 0 \leq y \leq 1$$

- 1). 基于样本写出对数似然函数; (5 分)
- 2). 假设 $n = 10$, $\sum_{i=1}^{10} \log Y_i = -5$, $\sum_{i=1}^{10} \log(1 - Y_i) = -10$ 。假设 Newton-Raphson 极大似然估计算法中未知参数 θ 的当前值为 $\theta^{(t)} = 1$, 请给出 $\theta^{(t+1)}$, 即, 推导出 $\theta^{(t+1)}$ 。(10 分)