

Statistics – Information Criterion

Yabusame

2024-04-28

1 信息准则的几个谣传

AIC 和 BIC 的原始形式如下:

$$\text{AIC}_m = -2 \sum_{i=1}^n \log(p_{\hat{\theta}}(y_i)) + 2d_m$$

$$\text{BIC}_m = -2 \sum_{i=1}^n \log(p_{\hat{\theta}}(y_i)) + d_m \log(n)$$

1.1 AIC 适合预测, BIC 适合解释 ?

这种看法忽略了参数化和非参数化情形, AIC 可能在 $n \rightarrow \infty$ 时也不能取到最好模型, 在非参数情况下, BIC 选择生成数据模型的一致性也不是良好定义的。

1.2 应该使用 AIC, 因为现实情况更常见到非参数化的情形 ?

在实际情况中信息准则也会受到样本数的影响, 例如:

1. 在样本少非参数的情况下, BIC 可以察觉到突出的模型
2. 在参数化的情况下, 系数在不同的数量级上很小, 并且样本数不足以估计它们, 在这种情况下选择模型使用 AIC 更加适合
3. 主要有部分参数化、部分非参数化的两种情况

1.3 penlity l_0 不如 penlity (LASSO, SCAD, MCP), 因为它是离散的 ?

信息准则相当于带有 l_0 penlity 的回归, penlity 是否正确取决于计算的目的

- 即使是对于固定调整参数(fixed tuning parameter), 选择模型的能力与 penlity function 的连续与否没有直接关系
- 固定调整参数(fixed tuning parameter)的选择基于数据, 其他 penlity 不一定会给出更好结果
- 事实上 l_0 penlity 方式以最少的约束条件得到了 minimax rate 最佳值

复杂的理论看不懂...直接到建议部分:

1. 当选择模型是为了预测的时候:
 - 依据参数化指标或者交叉验证选择 AIC-Type 或 BIC-Type 方法
 - 使用适应性的信息准则(例如 BC)结合 AIC 和 BIC 的方式
2. 当希望从后随观测的相似的样本大小中重新得到选择参数时(选择变量用以解释模型)—使用 BIC-Type 防止引入不重要的变量
3. 如果保护最坏情况预测精度非常重要, 首选 AIC — 极小极大速率(minimax rate)最优性 此外对于寻找可能相关变量的探索性研究, 尽管 AIC 型方法可能会过度选择, 但不会遗漏重要变量, 这些变量可以在大样本量的后续研究中验证

4. 当预测变量的数量 d 与样本大小 n 相比并不小，并且考虑 d 变量的所有子集时，最好使用高维 AIC 或 BIC 来解决潜在的严重选择偏差
5. 当模型选择不稳定性较高时，出于预测目的，可以考虑模型平均