

模型选择准则——AIC 和 BIC

吴思思 15320171151911 经济系西方经济学

2019 年 4 月 21 日

一、模型选择准则简介

二、Akaike information criterion,AIC

三、Bayesian information criterion,BIC

四、AIC 与 BIC 的比较

一、模型选择准则简介

模型选择是指从一组备选模型中选出与真实模型最为接近的一个模型，从而能够更好地拟合潜在的数据。模型选择问题通常包含模型类型选择和自变量选择两方面。此处，我们不涉及模型结构的选择而是围绕模型参数数量来探讨什么是合适的模型选择准则。通常存在很多备选的解釋变量可用于预测因变量，但没有必要在回归模型中包含所有的解釋变量。这里存在着权衡取舍：一方面，解釋变量越多，模型的系统偏差越小，如果所有的参数估计都不存在误差，那么该模型的预测能力也是最优的。但是另一方面，在随机样本容量 n 给定下，参数越多，参数估计的准确性越差。统计学上有一个重要的思想成为“KISS (keep it sophisticatedly simple)”原则，就是尽量用简单的模型刻画数据中包含的重要信息。下面介绍的两个常用的模型选择准则 AIC 和 BIC 都是这一思想的体现。

二、Akaike information criterion,AIC

Akaike information criterion, 简称 AIC, 是衡量统计模型拟合优良性的一种标准, 是由日本统计学家赤池弘次创立和发展的。该准则建立在熵的概念基础上, 可以权衡所估计模型的复杂度和此模型拟合数据的优良性。通常情况下, AIC 定义为:

$$AIC = -2 \ln(L) + 2k$$

其中 k 是模型参数个数, L 是似然函数。从一组可供选择的模型中选择最佳模型时, 通常选择 AIC 最小的模型。

当两个模型之间存在较大差异时, 差异主要体现在似然函数项。当似然函数差异不显著时, 上式第二项, 即模型复杂度则起作用, 从而参数个数少的模型是较好的选择。

一般而言, 当模型复杂度提高 (k 增大) 时, 似然函数 L 也会增大, 从而使 AIC 变小, 但是 k 过大时, 似然函数增速减缓, 导致 AIC 增大, 模型过于复杂容易造成过拟合现象。目标是选取 AIC 最小的模型, AIC 不仅要提高模型拟合度 (极大似然), 而且引入了惩罚项, 使模型参数尽可能少, 有助于降低过拟合的可能性。

线性回归模型可通过选择合适的解释变量数目 k , 以最小化下面的 Akaike 信息准则来选择模型:

$$AIC = \ln(s^2) + \frac{2k}{n}$$

其中 $s^2 = \frac{e'e}{n-k}$, k 是自变量的数目, 第一项 $\ln(s^2)$ 测度模型的拟合优度, 而第二项 $\frac{2k}{n}$ 测度模型的复杂程度。另外, s^2 是 $E(\varepsilon_t^2) = \sigma^2$ 的残差方差估计量 (residual variance estimator)。

三、Bayesian information criterion,BIC

BIC (Bayesian Information Criterion) 贝叶斯信息准则与 AIC 相似, 用于模型选择, 是 Schwartz 在 1978 年根据 Bayes 理论提出的判别准则。训练模型时增加参数数量, 也就是增加模型复杂度, 会增大似然函数, 但是也会导致过拟合现象, 针对该问题, AIC 和 BIC 均引入了与模型参数个数相关的惩罚项, BIC 的惩罚项比 AIC 的大, 考虑了样本数量, 样本数量过多时, 可有效防止模型精度过高造成的模型复杂度过高。通常情况下, BIC 定义为:

$$BIC = -2 \ln(L) + \ln(n) * k$$

其中 L 是在该模型下的最大似然, n 是数据数量, k 是模型的变量个数。 $\ln(n) * k$ 惩罚项在维数过大且训练样本数据相对较少的情况下, 可以有效避免出现维度灾难现象。

线性回归模型也可通过选择合适的 K, 以最小化下面的 Bayesian 信息准则来选择模型:

$$BIC = \ln(s^2) + \frac{k \ln(n)}{n}$$

AIC 和 BIC 两信息准则都试图在模型的拟合优度 $\ln(s^2)$ 与尽量少用参数之间进行权衡。当 $n \geq 7$ 时, $\ln n \geq 2$, 这时, 与 AIC 相比, BIC 对模型复杂度给予更大的惩罚。

四、AIC 与 BIC 的比较

AIC 与 BIC 的区别主要在于它们的构造方式不同，AIC 倾向于选择具有最优预测能力的模型。相对于 BIC 而言，AIC 选择的参数更多；而 BIC 倾向于选择准确的 K 值。在一定的正则条件下，BIC 可一致地选择正确的 K 值。从某种意义上来说，当 $n \rightarrow \infty$ 时，BIC 在大样本条件下更接近真实模型。而对 AIC 来说，不管样本容量 n 多大，它都会倾向于选择更多参数。在小样本下，上述描述不一定正确。在实践中，最优的 AIC 模型往往也接近于最优的 BIC 模型。通常，它们都给出同一最优模型。

参考文献

- [1]Advanced Econometrics,Yongmiao Hong
- [2]Harvard AM207,Understanding AIC
- [3]<https://blog.csdn.net/ljzology/article/details/81287500>
- [4] 百度百科-赤池信息量准则
- [5] 百度百科-贝叶斯信息准则