Consistency of LOOBIC

Zheng Yuan

November 2019

1 Model Setup

Given n observation pairs (x_i, y_i), $i = 1, 2, ..., n$, we consider the linear regression model

$$y_i = x_i'\beta + \epsilon_i,$$

where x_i's and β are p-dimensional vectors, ϵ_i's are iid random variables with mean 0 and variance σ^2.

Denote $X = (x_1, ..., x_n)'$ as the $n \times p$ dimensional design matrix, $y = (y_1, ..., y_n)'$ as the response vector and ϵ as the error vector, then the model can be written as

$$y = X\beta + \epsilon$$

.

2 Bayesian Information Criterion (BIC)

The BIC (Bayesian information criterion) was developed by Gideon E. Schwarz and published in a 1978 paper,

$$Choose\ the\ model\ for\ which\ logM_j(X_1, X_2, ..., X_n) - \frac{1}{2}k_j * logn\ is\ the\ largest$$

Later, the BIC is formally defined as

$$BIC(M_\alpha) = -2log(\hat{L}_\alpha) + d_\alpha logn$$

* $\hat{L}_\alpha$= the maximized value of the likelihood function of the model M_α, i.e., $\hat{L}_\alpha = p(X|\hat{\theta}_\alpha, M_\alpha)$, where $\hat{\theta}_\alpha$ are the parameter values that maximize the likelihood function

* X: the observed data

* n: the sample size

* d_α: the number of parameters estimated by model M_α

Under Gaussian linear models $M : Y = X\beta + \epsilon$,

$$\begin{aligned} logL(\hat{\beta}, \hat{\sigma}^2) &= log\Big\{(2\pi\hat{\sigma^2})^{-\frac{n}{2}} exp\Big\{-\frac{1}{2\hat{\sigma^2}}||Y - X\hat{\beta}||^2\Big\}\Big\} \\ &= -\frac{n}{2}log(2\pi\hat{\sigma^2}) - \frac{1}{2\hat{\sigma^2}}||Y - X\hat{\beta}||^2 \\ &= -\frac{n}{2}log(2\pi) - \frac{n}{2}log\frac{RSS}{n} - \frac{n}{2} \\ &= C - \frac{n}{2}log\frac{RSS}{n}. \end{aligned}$$

Therefore, BIC can be rewritten as

$$BIC(M_\alpha) = nlog\frac{RSS_\alpha}{n} + d_\alpha * logn = nlog(\hat{\sigma}^2_\alpha) + d_\alpha * logn, \tag{1}$$

up to an additive constant, which depends only on n and not on the model. The one with the lowest BIC is preferred when selecting from several models.

3 LOOBIC

By combining the LOOCV and BIC together, a new information criterion, LOOBIC is proposed and is defined as follows:

$$LOOBIC(M_\alpha) = n * log(\hat{\Gamma}^{CV}_{\alpha,n}) + d_\alpha * logn \tag{2}$$

Following the notation system developed in Shao (1993), $\hat{\Gamma}^{CV}_{\alpha,n}$ denotes the CV(1) estimate of true prediction error for M_α, that is, $\Gamma_{\alpha,n}$, in Gaussian linear models :

$$\begin{aligned} \hat{\Gamma}^{CV}_{\alpha,n} &= \frac{1}{n}\sum_{i}^{n}[(1 - p_{i\alpha})^{-1}(y_i - x^T_{i\alpha}\hat{\beta}_\alpha)]^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{(Y_i - \hat{Y_{i,\alpha}})^2}{(1 - p_{i\alpha})^2}, \end{aligned}$$

where $o_{i\alpha}$ is the ith diagonal entry in the projection matrix for M_α, $p_{i\alpha} = [X_\alpha(X^T_\alpha X_\alpha)^{-1}X^T_\alpha]_{ii}$.

LOOBIC automatically induces a model selection procedure itself. Specifically, the one with the lowest LOOBIC is preferred when selecting from several models.

4 Consistency of LOOBIC

As it is shown in **Theorem 1** in "Leave One Out Cross Validation" report:
If M_α is in Category II, then

$$\hat{\Gamma}^{CV}_{\alpha,n} = \frac{1}{n}e^T e - \frac{1}{n}e^T P_\alpha e + \frac{2}{n}d_\alpha \sigma^2 + o(n^{-1}). \tag{3}$$

As $n \to \infty$, we have $\frac{2}{n}d_\alpha \sigma^2 + o(n^{-1}) \to 0$

Also,

$$\begin{aligned}
\frac{1}{n}e^T e - \frac{1}{n}e^T P_\alpha e &= \frac{1}{n}e^T (I - P_\alpha)e \\
&= \frac{1}{n}(Y - X_\alpha \beta_\alpha)^T (I - P_\alpha)(Y - X_\alpha \beta_\alpha) \\
&= \frac{1}{n}Y^T (I - P_\alpha)Y - 2Y^T (I - P_\alpha)X_\alpha \beta_\alpha + \beta_\alpha^T X_\alpha^T (I - P_\alpha)X_\alpha \beta_\alpha \\
&= \frac{1}{n}Y^T (I - P_\alpha)Y.
\end{aligned}$$

because

$$(I - P_\alpha)X_\alpha \beta_\alpha = X_\alpha \beta_\alpha - X_\alpha (X_\alpha^T X_\alpha)^{-1} X_\alpha^T X_\alpha \beta_\alpha = X_\alpha \beta_\alpha - X_\alpha \beta_\alpha = 0$$

.

Furthermore,

$$\begin{aligned}
\frac{1}{n}e^T e - \frac{1}{n}e^T P_\alpha e &= \frac{1}{n}Y^T (I - P_\alpha)Y \\
&= \frac{1}{n}((I - P_\alpha)Y)^T (I - P_\alpha)Y \\
&= \frac{1}{n}(Y - X_\alpha \hat{\beta_\alpha})^T (Y - X_\alpha \hat{\beta_\alpha}) \\
&= \frac{RSS_\alpha}{n}
\end{aligned}$$

This shows that

$$\lim_{n \to 0}[\hat{\Gamma}^{CV}_{\alpha,n} - \frac{RSS_\alpha}{n}] = 0, \tag{4}$$

which implies that

$$\lim_{n \to 0}[log\hat{\Gamma}^{CV}_{\alpha,n} - log\frac{RSS_\alpha}{n}] = 0 \tag{5}$$

Therefore, it follows from (1)(2)(5) that

$$LOOBIC(M_\alpha) \to BIC(M_\alpha) \ as \ n \to \infty.$$

This completes the proof that LOOBIC is consistent when selecting models from Category II.