WBIC

渡辺澄夫（東京工業大学）

(1) 学習データ $D=(X_1, X_2, \ldots, X_n)$ は独立に確率分布 $q(x)\,dx$ に従う.

(2) $p(x|w)$ は w をパラメータとする学習モデル（統計モデル）.

(3) 事前分布 $\varphi(w)$.

(4) w_0 は $\int q(x)\log(q(x)/p(x|w))dx$ を最小にするパラメータ

(4) 対数損失関数 $L_n(w) = -\ (1/n) \sum_{i=1}^{n} \log p(X_i|w)$

(5) ベイズ自由エネルギー $F = -\log \int \prod_{i=1}^{n} p(X_i|w)\, \varphi(w)\, dw$

(6) $$\mathrm{WBIC} = \frac{\int n L_n(w) \prod_{i=1}^{n} p(X_i|w)^{\beta}\, \varphi(w)\, dw}{\int \prod_{i=1}^{n} p(X_i|w)^{\beta}\, \varphi(w)\, dw}$$ （ただし $\beta=1/\log n$ ）

(7) 実対数閾値 λ, 多重度 m

[定理] $q(x)$, $p(x|w)$, $\varphi(w)$ が何であっても下記が成立.

$$\boxed{F = nL_n(w_0) + \lambda \log n - (m-1)\log\log n + O_p(1)}$$

$$\boxed{\mathrm{WBIC} = nL_n(w_0) + \lambda \log n + O_p((\lambda \log n)^{1/2})}$$

(1) S. Watanabe, A widely applicable Bayesian information criterion, to appear in Journal of Machine Learning Research.