

Posterior Covariance Information Criterion

矢野 恵佑 数理・推論研究系 准教授

概要:

- MCMCを利用したモデル評価規準WAICを近年のデータ解析で現れる様々な予測状況に対応できるように拡張
- 提案規準の理論的性質を議論

背景:

- Widely-Applicable Information Criterion (WAIC; [1])は様々な分野で活用されている予測モデル評価規準
- 一回のMCMCのみで計算可能

WAIC: モデル $\{h(\cdot | \theta): \theta \in \Theta\}$ に対して以下を計算 (ただし $E_{\text{pos}}, V_{\text{pos}}$ は事後平均・事後分散)

$$\text{WAIC} = T_n + V_n$$

$$T_n := -\frac{1}{n} \sum_{i=1, \dots, n} \log E_{\text{pos}}[h(X_i | \theta)] \quad V_n := \frac{1}{n} \sum_{i=1, \dots, n} V_{\text{pos}}[\log h(X_i | \theta)]$$

特徴:

- ベイズ予測分布 $h_{\pi}(Y) := E_{\text{pos}}[h(Y | \theta)]$ を介してモデルの良さを評価
- 実際, WAICはベイズ予測分布の期待対数損失の漸近不偏推定量である $E[\text{WAIC}] = E[-\log E_{\text{pos}}[h(X_i | \theta)]] + o(1)$

様々な予測の状況:

- 近年のデータ解析では, 学習に利用する評価関数と予測に利用する評価関数が異なる予測状況が現れる
- 例: 共変量シフト, 因果推論, M-推定

Posterior Covariance Information Criterion:

- 学習時の評価関数 $\{s_i(X_i, \theta): \theta \in \Theta\}, i = 1, \dots, n$ / 予測時の評価関数 $\{h_i(X_i | \theta): \theta \in \Theta\}, i = 1, \dots, n$,
- 観測ごとの重み $w_i > 0, i = 1, \dots, n$
- 擬ベイズ事後標本 $\pi(\theta; X) \propto \exp(\sum_i s_i(X_i, \theta))$ に基づく平均 \bar{E}_{pos} ・共分散 $\overline{\text{Cov}}_{\text{pos}}$

を利用して以下を計算

$$\text{PCIC} = T_n + V_n$$

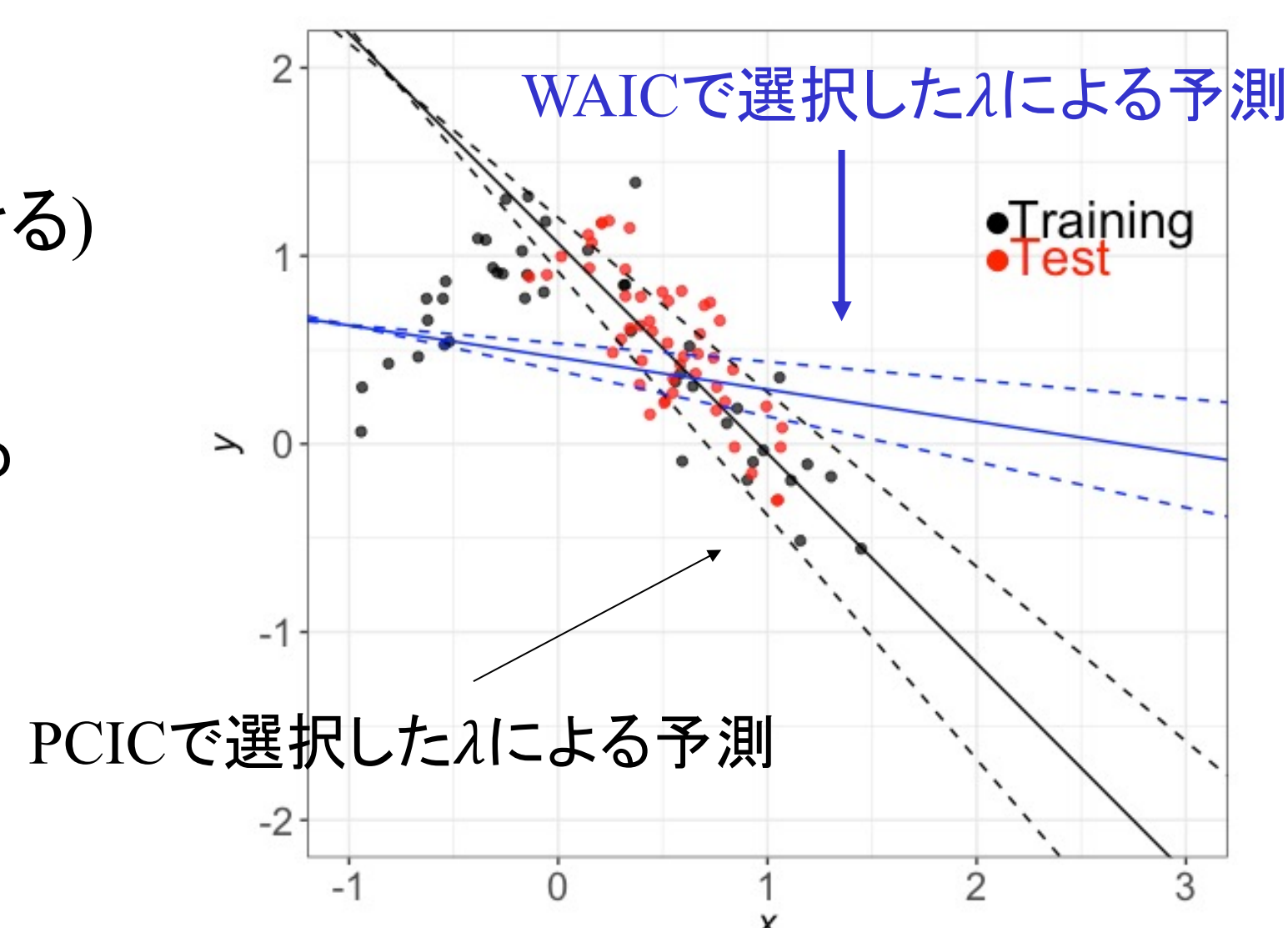
$$T_n := -\frac{1}{n} \sum_{i=1, \dots, n} w_i \log \bar{E}_{\text{pos}}[h_i(X_i | \theta)] \quad V_n := \frac{1}{n} \sum_{i=1, \dots, n} w_i \overline{\text{Cov}}_{\text{pos}}[\log h_i(X_i | \theta), s_i(X_i, \theta)]$$

理論的性質:

- PCICは擬ベイズ予測分布の重み付き期待対数損失の漸近不偏推定量
- 重み付きのGIC [2]と漸近等価になる (PCICは擬ベイズ標本のみで計算できる)

共変量シフトへの応用:

- 学習時の共変量の分布 $p_{\text{train}}(X)$ と予測時の共変量の分布 $p_{\text{test}}(X)$ が異なる
- $w_i = p_{\text{test}}(X_i)/p_{\text{train}}(X)$ による重み付き尤度による学習が有効 [3]
- $s_i(X_i, \theta) = w_i^\lambda \log h(X_i | \theta)$ として $\lambda > 0$ をWAIC, PCICによって選択
- 1次関数+ガウスノイズを $h(X_i | \theta)$ として利用
- PCICによる選択はテスト標本をよく予測



本研究は伊庭幸人教授(モデリング研究系)との共同研究である. WAICの性質・更なる一般化については伊庭教授のポスターを参照

参考文献:

- [1] S. Watanabe (2018). Mathematical Theory of Bayesian Statistics. Chapman & Hall/CRC.
- [2] S. Konishi & G. Kitagawa (1996). Generalised information criteria in model selection. *Biometrika*, 83, 875–890
- [3] H. Shimodaira (2000). Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90, 227–244.