ADVANCED BAYESIAN MODELING

Bayesian Cross-Validation

Recall evaluation of a fitted model: how well it predicts *new* data.

Recall problem with using deviance for model evaluation:

Same data y is used both for model fitting and for estimating the expected (logarithmic) score, which makes evaluation too optimistic.

This is why criteria like DIC include a correction for effective number of parameters.

Data Splitting

A different approach –

Split data y into two parts (preferably at random):

$$y \;=\; (y_{\text{train}},\, y_{\text{eval}})$$

Use y_{train} to fit the model and y_{eval} to estimate its predictive accuracy (score).

Issues:

- Depends on which split of data is used.
- Biased evaluation because fit is based on only a subset of the data set.

Cross-Validation

Cross-validation: Use many different data splits, and combine the evaluation results.

Leave-one-out cross-validation (LOO-CV): For n observations, use the n splits in which y_{eval} has only one observation.

This reduces evaluation bias, since y_{train} is almost the full data.

Bayesian Evaluation

For Bayesian models, generally assume y_{train} and y_{eval} are conditionally independent, then use log predictive density of y_{eval} based on y_{train}

$$\log p_{\text{train}}(y_{\text{eval}}) \quad = \quad \log \int p(y_{\text{eval}} \mid \theta)\, p_{\text{train}}(\theta)\, d\theta$$

where $p_{\text{train}}(\theta)$ is the posterior using only y_{train}.

In LOO-CV, each y_{eval} is just a y_i ($i = 1, \ldots, n$), and we denote the log predictive densities as

$$\log p_{\text{post}(-i)}(y_i) \qquad\qquad i = 1, \ldots, n$$

Bayesian LOO-CV

We choose the combined LOO-CV estimate to be the sum

$$\text{lppd}_{\textsf{loo-cv}} \;=\; \sum_{i=1}^{n} \log\, p_{\textsf{post}(-i)}(y_i)$$

for which larger values indicate better models.

Note: Resembles

$$\text{lppd} \;=\; \sum_{i=1}^{n} \log\, p_{\textsf{post}}(y_i)$$

used in WAIC.

Note: The Monte Carlo approximation to $\log p_{\mathsf{post}(-i)}(y_i)$ is

$$\log\left(\frac{1}{S}\sum_{s=1}^{S} p(y_i \mid \theta^{is})\right)$$

where

$$\theta^{i1}, \ldots, \theta^{iS}$$

is a sample from the posterior when y_i is left out.

Bayesian LOO-CV needs a total of n such samples: expensive for large n.

Remarks (BDA3, Sec. 7.2):

- There is a correction for the remaining bias in LOO-CV (usually small if n is large).
- There is a way to define an effective number of parameters.
- Bayesian LOO-CV is asymptotically equivalent to WAIC (under some conditions).