Bayesian Model Averaging

Donald E. Brown

School of Data Science
University of Virginia
Charlottesville, VA 22904

Averaging and Ensembles

- Rather than select, combine models
- Ensembles have demonstrated advantages
  - Improved accuracy
  - Reduced variance
  - Explicitly account for uncertainty in the choice of model
- Problems
  - Interpretability
  - Choice of method for combination
  - Cost to build

Bayesian Model Averaging

- Let y be a target variable, $\mathcal{D}$ be the data, and $M_k, k = 1, \ldots, K$ be the models with parameters $\boldsymbol{\theta}$.
- Then an averaged estimate for the target is

$$p(y|\mathcal{D}) = \sum_{k=1}^{K} p(y|M_k, \mathcal{D}) p(M_k|\mathcal{D})$$

- The posterior probability for a model, M_k is

$$p(M_k|\mathcal{D}) = \frac{p(\mathcal{D}|M_k)p(M_k)}{\sum_{i=1}^{I} p(\mathcal{D}|M_i)p(M_i)}$$

- The likelihood, $p(\mathcal{D}|M_k)$ is

$$p(\mathcal{D}|M_k) = \int p(\mathcal{D}|M_k, \boldsymbol{\theta}) p(\boldsymbol{\theta}|M_k) d\boldsymbol{\theta}$$

Pseudo -Bayesian Model Averaging

- Rather than weight from the model posterior, use weights from information or other criteria
- WAIC weights: Let $\Delta_k = \text{WAIC}_k - \text{WAIC}^*$ where WAIC^* is the lowest WAIC score. Then use weights:

$$w_k = \frac{exp(-\frac{1}{2}\Delta_k)}{\sum_{i=1}^{K} exp(-\frac{1}{2}\Delta_i)}$$

- Then the BMA is

$$p(y|\mathcal{D}) = \sum_{k=1}^{K} p(y|M_k, \mathcal{D}) w_k$$