Bayesian data analysis – reading instructions 7

Aki Vehtari

Chapter 7

Outline of the chapter 7

- 7.1 Measures of predictive accuracy
- 7.2 Information criteria and cross-validation (read instead the article mentioned below)
- 7.3 Model comparison based on predictive performance (read instead the article mentioned below)
- 7.4 Model comparison using Bayes factors
- 7.5 Continuous model expansion / sensitivity analysis
- 7.5 Example (may be skipped)

Instead of Sections 7.2 and 7.3 it's better to read

- Aki Vehtari, Andrew Gelman and Jonah Gabry (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. In Statistics and Computing, 27(5):1413-1432, doi:10.1007/s11222-016-9696-4. arXiv preprint arXiv:1507.04544.
- LOO package glossary summarises many important terms used in the assignments.

In Sections 7.2 and 7.3 of BDA, for historical reasons there is a multiplier -2 used. After the book was published, we have concluded that it causes too much confusion and recommend not to multiply by -2. The above paper is not using -2 anymore.

See also extra material at https://avehtari.github.io/modelselection/

- Videos, slides, notebooks, references
- The most relevant for the course is the first part of the talk "Model assessment, comparison and selection at Master class in Bayesian statistics, CIRM, Marseille"
- Sections 1 and 5 (less than 3 pages) of "Uncertainty in Bayesian Leave-One-Out Cross-Validation Based Model Comparison" clarify how to interpret standard error in model comparison
- Cross-validation FAQ https://avehtari.github.io/modelselection/CV-FAQ.html answers many frequently asked questions.

Find all the terms and symbols listed below. When reading the chapter and the above mentioned article, write down questions related to things unclear for you or things you think might be unclear for others.

- predictive accuracy/fit/error
- external validation
- cross-validation
- information criteria
- overfitting
- measures of predictive accuracy
- point prediction
- scoring function

- mean squared error
- probabilistic prediction
- scoring rule
- logarithmic score
- log-predictive density
- out-of-sample predictive fit
- elpd, elppd, lppd
- deviance
- within-sample predictive accuracy
- adjusted within-sample predictive accuracy
- AIC, DIC, WAIC (less important)
- effective number of parameters
- singular model
- BIC (less important)
- leave-one-out cross-validation
- evaluating predictive error comparisons
- bias induced by model selection
- Bayes factors
- continuous model expansion
- sensitivity analysis

Additional reading

More theoretical details can be found in

- Aki Vehtari and Janne Ojanen (2012). A survey of Bayesian predictive methods for model assessment, selection and comparison. In Statistics Surveys, 6:142-228. http://dx.doi.org/10.1214/12-SS102

See more experimental comparisons in

- Juho Piironen and Aki Vehtari (2017). Comparison of Bayesian predictive methods for model selection. Statistics and Computing, 27(3):711-735. doi:10.1007/s11222-016-9649-y. http://link.springer.com/article/10.1007/s11222-016-9649-y

Posterior probability of the model vs. predictive performance

Gelman: "To take a historical example, I don't find it useful, from a statistical perspective, to say that in 1850, say, our posterior probability that Newton's laws were true was 99%, then in 1900 it was 50%, then by 1920, it was 0.01% or whatever. I'd rather say that Newton's laws were a good fit to the available data and prior information back in 1850, but then as more data and a clearer understanding became available, people focused on areas of lack of fit in order to improve the model."

Newton's laws are still sufficient for prediction in specific contexts (non-relative speeds and differences in gravity, non-significant effects of air resistance or other friction). See more in the course video 1.1 Introduction to uncertainty and modelling https://aalto.cloud.panopto.eu/Panopto/Pages/Viewer.aspx?id=d841f429-9c3d-4d24-8228-a9f400efda7b.