ADVANCED BAYESIAN STATISTICAL LEARNING 2023 - ASU

Professor: Hedibert Freitas Lopes

Syllabus

Course description: The end of the course goal is to expose the student to modern Bayesian solutions to highly structured and stochastic real world problems. We will visit well known Bayesian issues, such as prior specification/sensitivity, model comparison/criticism and model averaging, as well as Bayesian computation via various Monte Carlo methods. We approach regularization in linear and log-linear models via Bayesian LASSO, Spike-and-Slab priors and related sparsity-inducing priors. We cover decoupling shrinkage and selection strategies in a fully Bayesian decision framework. Other topics covered are finite and infinite mixtures for Bayesian semi- and non-parametric modeling, large-scale (dynamic/spatial) factor models, Bayesian additive regression trees (BART), Bayesian text modeling and modeling large-scale time-varying covariance matrices. All classroom examples and implementations as well as projects will be carried out by the open-source statistical software R.

Useful textbooks:

- Gamerman and Lopes (2006) MCMC: Stochastic Simulation for Bayesian Inference, 2nd Edition.
- Gelman, Carlin, Stern, Dunson, Vehtari and Rubin (2020) Bayesian Data Analysis, 3rd Edition.
- Migon, Gamerman and Louzada (2015) Statistical Inference: An Integrated Approach, Second Edition.
- Hoff (2009) A First Course in Bayesian Statistical Methods. Springer.

Homework assignments: HW1 + HW2 + HW3 + HW4

List of papers for final presentation

- The illusion of the illusion of sparsity – Fava and Lopes (2021), Brazilian Journal of Probability and Statistics, 35(4), 699-720 https://arxiv.org/abs/2009.14296
- A weakly informative default prior distribution for logistic and other regression models – Gelman, Jakulin, Pittau and Zu (2008), Annals of Applied Statistics, 2(4), 1360-1383. https://doi.org/10.1214/08-AOAS191
- Do forecasts of bankruptcy cause bankruptcy? A machine learning sensitivity analysis – Papakostas, Hahn, Murray, Zhou and Gerakos (2023), Annals of Applied Statistics, 17(1), 711-739. https://doi.org/10.1214/22-AOAS1648
- Forecasting with many predictors using Bayesian additive regression trees – Pruser (2019), Journal of Forecasting, Volume38, Issue7, November 2019, Pages 621-631. https://doi.org/10.1002/for.2587.

TEACHING MATERIAL

Bayesian ingredients

1. Basic Bayes
   - R code: Bernoulli iid, logit regression or probit regression? (Rmarkdown)
   - R code: Computing coronavac efficacy with logit, probit and clog-log generalized linear models?
   - R code: Space Shuttle Challenger O-Ring – binary/glm binomial
   - Count data: Poisson model, Gamma prior vs Poisson model, Log-normal prior (R code)
2. Exchangeability
3. Principles of data reduction
   - Discussion about p-values – P-values not only violate conditionality principle, but it is commonly mistaken as “the probability that the null hypothesis is true”. Recall that, Pr(H0 is true|data) is a well-defined Bayesian quantity, while the p-value is the probability of the data (or its more extreme versions) given that the null hypothesis is true: Pr(data|H0 is true); a totally different quantity!
4. Decision theory + More on estimators
   - For those keen to learn a bit more about Bayesian statistical decision theory beyond my meager lecture notes, I recommend a few places: 1) Statistical Decision Theory and Bayesian Analysis (2nd edition) – Berger (1985); 2) The Bayesian Choice (2nd edition) – Robert (2007); 3) Decision Theory: Principles and Approaches – Parmigiani & Inoue (2009); 4) Lecture notes on “Bayes Methods and Elementary Decision Theory” by Wellner (University of Washington); and Lecture notes on “Evaluating the performance of estimators” by Pati (Texas A&M University).
5. Bayesian model criticism
6. Additional reading material:
   - Chapter 2 of Gamerman and Lopes (2006) – Compact, but easy to read.
   - Chapters 2-4 of Migon, Gamerman and Louzada (2014) – Classical and Bayesian inference.
   - Chapter 1 and 2 of Gelman et al. (2013) – Application-oriented.
   - Chapter 4 (Sections 4.1-4.4) of Berger (1985) – More technical

Bayesian Computation

1. Monte Carlo (MC) Methods
   - *Monte Carlo Integration and MCI via Importance function*
   - *More on MCI*
   - *Bayesian multinomial model – MC integration & SIR*
   - *Gaussian vs Cauchy model & Gaussian prior*
   - *Bayesian binomial vs Poisson models – SIR (Analytical derivations)*
   - *Gaussian, Student's t and switching – MCI via Importance function for model comparison*
   - SIR, scale mixture of normals and raoblackwellization
2. Markov chain Monte Carlo (MCMC) algorithms
   - Markov chain: a brief review
   - *Random walk Metropolis: two toy examples*
   - *An example of linear regression with missing regressors*
   - *Linear regression with autocorrelated errors: SIR vs Gibbs*
   - *Nonlinear regression – random walk Metropolis*
   - *Hamiltonian Monte Carlo: A toy example*
   - *Stan/rstan for posterior inference: Hamiltonian MC (HMC) methods*
   - *A few more examples developed in class (Stan code)*
   - MC and MCMC: Key References

Bayesian Learning

1. Multiple linear regression: selection, shrinkage, sparsity
   - Motorcycle example
   - Slides from the 2015 School of Time Series and Econometrics tutorial
   - Hahn, He and Lopes (2018) Gaussian linear regression with arbitrary sparsity (slides of a talk)
   - Fava and Lopes (2021) The illusion of the illusion of sparsity (slides of a talk)
   - Michael Betancourt's Bayes Sparse Regression (stan/rstan example)
   - Example 1: Normal linear models: subset selection via BIC and shrinkage/sparsity via regularizing priors (data)
   - Example 2: Normal linear models: k-fold cross-validation (data + R code)
   - Example 3: More on linear models: In-sample vs out-of-sample MSE/MAE (data + R code)
   - Example 4: More on linear models: Orthogonal vs factor-based X
   - Example 5: More on linear models: Larger wage dataset
2. Classification: logistic regression and discriminant analysis
   - Sparse logistic regression for the spam/ham dataset (data)
   - Marketing campaigns of a Portuguese banking institution (data)
   - Sparse logistic regression: comparison of regularization and Bayesian implementations
   - Gelman, Jakulin, Pittau and Zu (2008) A weakly informative default prior distribution for logistic and other regression models, *AOAS*, 2(4), 1360-1383.
   - Polson, Scott and Windle (2013) Bayesian Inference for logistic models using Pólya-Gamma latent variables, *JASA*, 108, 1339-1349.
3. Bayesian factor analysis (BFA)
   - Bayesian factor analysis in R: a simple illustration (additional routines)
   - Factor models: an annotated bibliography (2003)
   - Lopes (2014)
   - Factor modeling of pollutant NO3 (Data)
4. Principal components analysis (PCA), PCA-based and FA-based regressions
5. Finite mixture of distributions
6. Spatial models
7. Bayesian CART
   - CART step-by-step for the Boston housing data
   - CART for the motorcycle data
   - Bootstrap illustration
8. Random forests
9. Bayesian additive regression trees (BART)
   - Example 1: ICU data: CART, BART and random forest (R code)
   - Example 2: Stock and Watson's (2002) macro data (data)
   - More examples: Four BART applications & 2 reviews + cute CART trees and 3D plots
   - More recent references
10. Latent Dirichlet Allocation (LDA)
11. Neural Networks
    - Bayesian two layer neural network (R code + tutorial in brnn package)
    - More comparisons (R code)
    -

Old homework assignments (spring 2022): *HW1* (*Solution*) + *HW2* (*Solution*) + *HW3* (*Derivations* + *R code*) + *HW4* (*Solution*)

Pool of papers from final presentation (spring 2022)