AMS 268: Advanced Bayesian Computation (Winter 2018) Course Desciption

January 6, 2018

Course description: This course is divided broadly into three sections:

(i) *High dimensional linear regression*: Penalized optimization methods, high dimensional Bayesian regression with spike-and-slab prior, g-prior, Bayesian compression. Multivariate high dimensional linear regression techniques are also discussed.

(ii) *Nonparametric regression models:* CART, Random Forest, BART, Gaussian process regression.

(iii) *Nonparametric regression models with large datasets:* Compressed Gaussian process, predictive process, compactly supported correlation functions, kernel convolutions.

Background: Prerequisite for this course is AMS 207 or equivalent.

Course objectives: This course is directed to provide knowledge on a few advanced Bayesian modeling techniques motivated by data applications from fields such as genomics, neuroscience, finance, economics and environmental science. The major objective of this course is to make an understanding of the widely used methodologies in machine learning and statistics and try them out as part of the assignments and class projects.

Textbook/course material: Most of the materials covered in this course are taken from research papers on big data in Bayesian statistics which have been published in the last decade. Therefore, research papers will be used as course materials mostly. Some of the course materials will be taken from the book:
Robert, C. and Casella, G. (2004). Monte Carlo Statistical Methods (Second Edition). New

York, Springer.

Relevant articles: Ideas in the following research papers will be discussed in detail.

1. Ishwaran, H. and Rao, S. (2005), "Spike and Slab Variable Selection: Frequentist and Bayesian Strategies," *Annals of Statistics*, **33**, 730–773.

2. George, E. I. and McCulloch, R. E. (1993), "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, **88**, 881-889.

3. Rockova, V. and George, E. I. (2015), "The Spike and Slab Lasso," `http://faculty.chicagobooth.edu/workshops/econometrics/PDF%202016/RockovaJMP.pdf`.

4. Liang, F., Paulo, R., Molina, G. Clyde, M.A. and Berger, J. O. (2008), "Mixture of g Priors for Bayesian Variable Selection," *Journal of the American Statistical Association*, **103**.

5. Kass, R. E. and Raftery, A. E. (1995), "Bayes Factor," *Journal of the American Statistical Association*, **90**, 773–795.

6. Guhaniyogi, R. and Dunson, D. B. (2016), "Bayesian Compressed Regression," *Journal of the American Statistical Association*, **512**, 1500-1514.

7. Guhaniyogi, R. and Dunson, D. B. (2016), "Compressed Gaussian Process for Manifold Regression," *Journal of the Machine Learning Research*, **17**, 1-26.

8. Brown, P. J., Vannucci, M. and Fearn, T. (1998), "Multivariate Bayesian Variable Selection and Prediction," *Journal of the Royal Statistical Society Series B*, **60**, 627–641.

A Detailed Course Description:
week 1: Penalized Optimization, ridge regression, lasso and its variants.
week 2: Bayesian variable selection, spike and slab prior.
Homework 1
week 3: Developments in spike and slab prior, developments in g-prior.
week 4: Paradoxes with g prior, mixtures of g prior.
Homework 2
week 5: Bayesian compressed regression, multivariate variable selection methods.
week 6: Classification and regression tree (CART), random forest (RF) and Bayesian additive regression tree (BART).
Homework 3

week 7: Gaussian process regression, high dimensional Gaussian processes, compressed Gaussian process.
week 8: Variable selection in Gaussian process with Automatic Relevance Determination (ARD), Gaussian processes with big data tacked using kernel convolution.
Homework 4
week 9: Predictive process, compactly supported correlation function, meta kriging.
week 10: Final presentations.

Homework: Three homework will be assigned (on a biweekly basis), and will be due two weeks later. Homework problems are assigned so that students code (or use available packages in R or Matlab) the methods or algorithms taught in the class.

End Term Project: End term project will be based upon replicating an already existing paper in this area of study. Any other idea on final projects is also appreciated. Each student has to submit a brief (4-5 pages) report containing his/her analysis and has to deliver a 15 minutes presentation highlighting his/her findings in the last week.

Course grade: Homework: 60%; End Term Project: 40%.