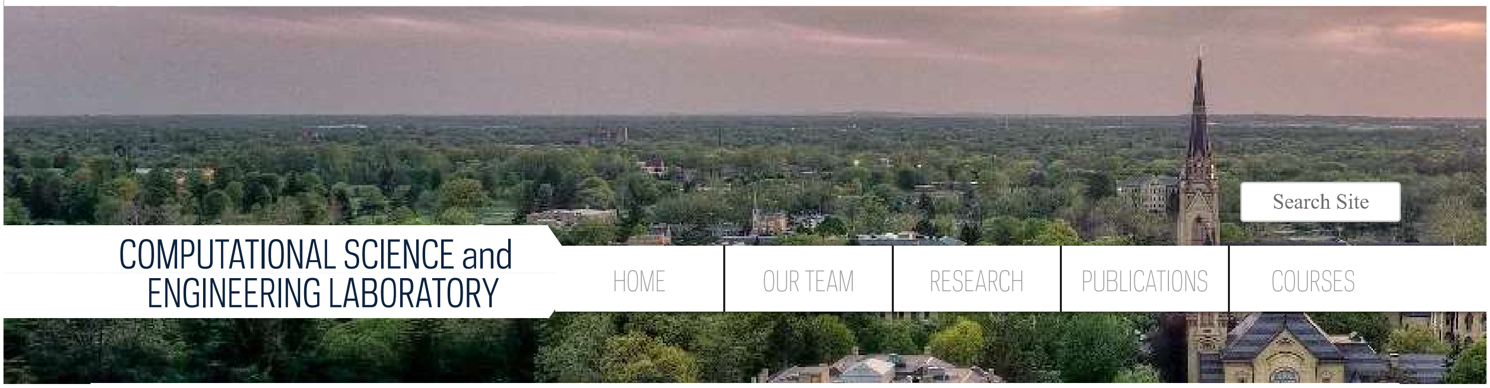
UNIVERSITY *of* NOTRE DAME COLLEGE *of* ENGINEERING



Bayesian Methods for Surrogate Modeling and Dimensionality Reduction

University of Notre Dame, Spring 2019

T R - 12:30A - 1:45 PM (Lectures, DeBartolo 129)

Friday: 11:30A - 12:20 PM (Recitation, DeBartolo 102)

Professor Nicholas Zabaras

Lecture notes and videos

1. Introduction to Machine Learning
    - Supervised and unsupervised learning, reinforcement learning; Regression and classification, Probabilistic predictions and point estimates; Examples, Document classification, Iris flower dataset, Image classification, Face detection and recognition; Supervised versus unsupervised learning; Unsupervised learning, Hidden/Latent variables, Discovering clusters, Dimensionality Reduction, Discovering graph structure, Matrix Completion; Parametric Vs. non parametric models; K-Nearest Neighbor Classifiers; Linear and Logistic Regression; Model selection, cross-validation and overfitting; The curse of dimensionality; No free lunch theorem.
      [Video-Lecture] [Lecture Notes]
2. Generative Bayesian Models for Discrete Data
    - Generative Models; Bayesian concept learning, Likelihood, Prior, Posterior, Posterior predictive distribution, Plug-in Approximation; The beta-binomial model, Likelihood, Prior, Posterior, Posterior predictive distribution, Blackswan paradoxes and Plug-in approximations, Outcome of multiple future trials, Beta-Binomial Distribution; The Dirichlet-multinomial model, Likelihood, Prior, Posterior, Posterior predictive, Bayesian Analysis of the Uniform Distribution, Language Model using Bag of Words; Naive Bayes classifiers, Examples, MLE for Naïve Bayes Classifier, Example for bag-of-words binary class model, Summary of the Algorithm, Bayesian Naïve Bayes, Using the model for prediction, The log-sum-exp trick, Feature selection using mutual information; Classifying documents using bag of words.
      [Video-Lecture] [Lecture Notes]
3. Generative Bayesian Models for Discrete Data (continued)
    - Generative Models; Bayesian concept learning, Likelihood, Prior, Posterior, Posterior predictive distribution, Plug-in Approximation; The beta-binomial model, Likelihood, Prior, Posterior, Posterior predictive distribution, Blackswan paradoxes and Plug-in approximations, Outcome of multiple future trials, Beta-Binomial Distribution; The Dirichlet-multinomial model, Likelihood, Prior, Posterior, Posterior predictive, Bayesian Analysis of the Uniform Distribution, Language Model using Bag of Words; Naive Bayes classifiers, Examples, MLE for Naïve Bayes Classifier, Example for bag-of-words binary class model, Summary of the Algorithm, Bayesian Naïve Bayes, Using the model for prediction, The log-sum-exp trick, Feature selection using mutual information; Classifying documents using bag of words.
      [Video-Lecture] [Lecture Notes]
4. Course Summary
    - Etc. .
      [Video-Lecture] [Lecture Notes]

Homework

- Sept. 13, Homework 1
  - Working with multivariate Gaussians, exponential family distributions, posterior for (μ,σ^2) for a Gaussian likelihood with conjugate prior, Bayesian information criterion (BIC), whitening vs standardizing the data.
    [Homework] [Solution] [Software]
- Sept. 20, Homework 2
  - Bla Bla
    [Homework] [Solution] [Software]
- Nov. 27, Homework 6
  - Principal Component Analysis, Bayesian PCA, EM Algorithm for PCA, Expectation Maximization, Gaussian Process Modeling.
    [Homework] [Solution] [Software]

Course info and references

Credit: 4 Units

Lectures: Tuesdays and Thursdays 12:30 -- 1:45 pm, DeBartolo Hall 129.

Recitation: Fridays. 11:30 -- 12:15 pm, DeBartolo Hall 101.

Professor: Nicholas Zabaras, 311 I Cushing Hall, `nzabaras@gmail.com`

Teaching Fellow (Volunteer): Dr. Souvik Chakraborty, `csouvik41@gmail.com`

Office hours: Mond. & Wedn. 5:00 -- 6:00 p.m., 311I Cushing.

Course description: The course covers selective topics on Bayesian scientific computing relevant to high-dimensional data-driven engineering and scientific applications. An overview of Bayesian computational statistics methods will be provided including Monte Carlo methods, exploration of posterior distributions, model selection and validation, MCMC and Sequential MC methods and inference in probabilistic graphical models. Bayesian techniques for building surrogate models of expensive computer codes will be introduced including regression methods for uncertainty quantification, Gaussian process modeling and others. The course will demonstrate these techniques with a variety of scientific and engineering applications including among others inverse problems, dynamical system identification, tracking and control, uncertainty quantification of complex multiscale systems, physical modeling in random media, and optimization/design in the presence of uncertainties. The students will be encouraged to integrate the course tools with their own research topics.

Intended audience: Graduate Students in Mathematics/Statistics, Computer Science, Engineering, Physical/Chemical/Biological/Life Sciences.

References of General Interest: The course lectures will become available on the course web site. For in depth study, a list of articles and book chapters from the current literature will also be provided to enhance the material of the lectures. There is no required text for this course. Some important books that can be used for general background reading in the subject areas of the course include the following:

References on (Bayesian) Machine Learning:

- C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2007.
- Kevin P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012 (a free ebook is also available from the author's web site).
- C. E. Rasmussen & C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006 (a free ebook is also available from the Gaussian Processes web site).

Homework: assigned every three to four lectures. Most of the homework will require implementation and application of algorithms discussed in class. We anticipate between five to seven homework sets. All homework solutions and affiliated computer programs should be mailed by midnight of the due date to `this Email address`. All attachments should arrive on an appropriately named zipped directory (e.g. HW1_Submission_YourName.rar). We would prefer typed homework (include in your submission all original files e.g. Latex and a Readme file for compiling and testing your software).

Term project: A project is required in mathematical or computational aspects of the course. Students are encouraged to investigate aspects of Bayesian computing relevant to their own research. A short written report (in the format of NIPS papers) is required as well as a presentation. Project presentations will be given at the end of the semester as part of a day or two long symposium.

Grading: Homework 60% and Project 40%.

Prerequisites: Linear Algebra, Probability theory, Introduction to Statistics and Programming (any language). The course will require significant effort especially from those not familiar with computational statistics. It is a course intended for those that value the role of Bayesian inference and machine learning on their research.

Syllabus

1. Review of probability and statistics
   - Laws of probability, Bayes' Theorem, Independency, Covariance, Correlation, Conditional probability, Random variables, Moments
   - Markov and Chebyshev Inequalities, transformation of PDFs, Central Limit Theorem, Law of Large Numbers
   - Parametric and non-parametric estimation
   - Operations on Multivariate Gaussians, computing marginals and conditional distributions, curse of dimensionality
2. Introduction to Bayesian Statistics
   - Bayes' rule, estimators and loss functions

   - Bayes' rule, estimators and loss functions
   - Bayes' factors, prior/likelihood & posterior distributions
   - Density estimation methods
   - Bayesian model validation.
3. Introduction to Monte Carlo Methods
   - Importance sampling
   - Variance reduction techniques
4. Markov Chain Monte Carlo Methods
   - Metropolis-Hastings
   - Gibbs sampling
   - Hybrid algorithms
   - Trans-dimensional sampling
5. Sequential Monte Carlo Methods and applications
   - Target tracking/recognition
   - Estimation of signals under uncertainty
   - Inverse problems
   - Optimization (simulated annealing)
6. Uncertainty Quantification Methods
   - Regression models in high dimensions
   - Gaussian process modeling
   - Forward uncertainty propagation
   - Uncertainty propagation in time dependent problems
   - Bayesian surrogate models
   - Inverse/Design uncertainty characterization
   - Optimization and optimal control problems under uncertainty
7. Uncertainty Quantification using Graph Theoretic Approaches
   - Data-driven multiscale modeling
   - Nonparametric Bayesian formulations

CSE Laboratory 372 Fitzpatrick Hall, Notre Dame, IN 46556, USA
Phone 574-631-2429 nzabaras@nd.edu
Contact us
Accessibility Information