Macao Polytechnic University

Faculty of Applied Sciences

Master of Science in Big Data and Internet of Things

Module Outline

Academic Year 2022/2023 Semester 2

Learning Module	Selected Topics II			**Class Code**	COMP6106
Pre-requisite(s)	Nil				
Medium of Instruction	English			**Credit**	3
Lecture Hours	45 hrs	**Lab/Practice Hours**	0 hrs	**Total Hours**	45 hrs
Instructor	Dr. Charles Lam Chi Kin		**E-mail**	cklamsta@mpu.edu.mo	
Office	N46B, Wui Chi Building, Main Campus		**Telephone**	8599-6823	

Description

This aim of this module is to provide students with an understanding of Bayesian statistics and to build students' ability to develop Bayesian models for practical data analysis problems. Students will learn to implement Bayesian models with Markov chain Monte Carlo and other numerical methods in software (e.g., R) and interpret the results. In addition, they will learn about the Bayesian perspective and its underlying theory.

Learning Outcomes

After completing the learning module, students will be able to:

1. Understand the basic notions of Bayesian statistics; (SM1p, EA1p)
2. Prove and use Bayes Theorem in its various forms; (SM2p, EA2p)
3. Explain the operation and basic theory of the two main Markov-Chain Monte-Carlo methods, Gibbs sampling and the Metropolis-Hastings algorithm; (SM2p, EA2p)
4. Derive the full conditional distributions for parameters in simple low-dimensional problems; (SM2p, EA2p)
5. Describe Bayesian approaches to model selection; (SM2p, EA2p)

6. Implement Gibbs sampling and the Metropolis-Hastings algorithm in R; (SM3p, EA3p)
7. Carry out a full Bayesian data analysis of a real data set by implementing MCMC methods and summarise their analysis and conclusions. (SM3p, EA4p)

Content

1. Introduction (4.5 hours)
   - 1.1 Belief functions and probabilities
   - 1.2 Events, partitions and Bayes' rule
   - 1.3 Independence, random variables, joint distributions
   - 1.4 Introduction tor R programming
2. Single-parameter models (4.5 hours)
   - 2.1 The binomial model
   - 2.2 The Poisson model
   - 2.3 Exponential families and conjugate priors
3. Monte Carlo approximation (4.5 hours)
   - 3.1 The Monte Carlo method
   - 3.2 Posterior inference for arbitrary functions
   - 3.3 Sampling from predictive distributions
4. The normal model (4.5 hours)
   - 4.1 The normal model
   - 4.2 Inference for the mean, conditional on the variance
   - 4.3 Joint inference for the mean and variance
   - 4.4 Bias, variance and mean squared error
5. Posterior approximation with the Gibbs sampler (6.0 hours)
   - 5.1 A semiconjugate prior distribution
   - 5.2 Discrete approximations
   - 5.3 Sampling from the conditional distributions
   - 5.4 Gibbs sampling
   - 5.5 Introduction to MCMC diagnostics
6. The multivariate normal model (3.0 hours)
   - 6.1 The multivariate normal density
   - 6.2 A semiconjugate prior distribution for the mean
   - 6.3 The inverse-Wishart distribution
   - 6.4 Gibbs sampling of the mean and covariance
7. Group comparisons and hierarchical modeling (4.5 hours)
   - 7.1 Comparing two groups
   - 7.2 Comparing multiple groups

7.3 The hierarchical normal model

8. Linear regression (4.5 hours)
   8.1 The linear regression model
   8.2 Bayesian estimation for a regression model
   8.3 Model selection

9. Nonconjugate priors and Metropolis—Hastings algorithms (4.5 hours)
   9.1 Generalized linear models
   9.2 The Metropolis algorithm
   9.3 Metropolis, Metropolis—Hastings and Gibbs

10. Linear and generalized linear mixed effects models (4.5 hours)
   10.1 A hierarchical regression model
   10.2 Full conditional distributions
   10.3 Generalized linear mixed effects models

Teaching Method

Lectures and tutorials.

Attendance

Attendance requirements are governed by the "Academic Regulations Governing Bachelor's Degree Programmes" of Macao Polytechnic University. Students who do not meet the attendance requirements for the module will not be permitted to sit the final or re-sit examination and shall be awarded an 'F' grade.

Assessment

This learning module is graded on a 100 point scale, with 100 being the highest possible score and 50 being the passing score.

	Item	Description	AHEP3 LO	Percentage
1.	Assignments / Classwork	Home- / Classroom-based exercises	SM2p, EA3p, D4p	20%
2.	Tests	Knowledge assessment	SM2p, EA3p, D4p	30%
3.	Examination	3-hour written examination	SM2p, EA3p, D4p	50%
			Total Percentage:	100%

Students with an overall score of less than 35 in the coursework will fail the module even if the overall score for the module is 50 or above.
Students with a score of less than 35 in the final examination will fail the module even if the overall score for the module is 50 or above.

Teaching Material(s)

Textbook(s)

There is no official text for this module. Module notes are distributed in the class.

Reference

Reference book(s)

1. Hoff, P.D. (2009). *A First Couse in Bayesian Statistical Methods*, Springer.
2. Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis* (3rd Edition). Boca Raton: CRC press.