CAS MA 751 – Advanced Statistical Methods II

Boston University, Spring 2010

Instructor
Professor Eric Kolaczyk
Office: MCS 223
Phone: 617-353-5208
Email: *kolaczyk@math.bu.edu*
Course Website: *http://math.bu.edu/people/kolaczyk/ma751.html*
Office Hours: Th2-3:30; F9:30-11.

Class Meetings: MWF 11am – noon, MCS B23.

Course Description: This course is nominally the second in a two-semester PhD sequence on post-classical statistical methods and their applications, but it may be taken as a stand-alone course. We will cover a selection of topics in statistical learning, such as regularized basis methods, kernel methods, boosting, neural networks, support vector machines, and graphical models.

Prerequisites: CAS MA 581 (Probability) and MA 575 (Linear Models); or consent of instructor.

Text:

T. Hastie, R. Tibshirani, and J. Friedman. *Elements of Statistical Learning: Data Mining, Inference, and Prediction (Second Edition).* Springer; New York. 2009.

Software: Instruction will be done using the statistical software **R** for computing in this course. The software is free and compatible with Windows, Mac, and Linux/Unix. It may be downloaded from `cran.us.r-project.org`. The main website for the **R** project is `www.r-project.org`. For those who do not wish to install **R** on their own machines, it can be accessed in a command-line version on the ACS machine 'acs-linux' and, for those students in the department, on the machines in the Department of Mathematics and Statistics.

Use of **R** is not required for the course. If you prefer to use a different package or programming language/environment, you are free to do so. But you will be responsible for seeing that you have sufficient access to software tools for the various topics covered in the course.

Course Format:

Much of the material in this course involves developments in the field of statistical/machine learning from just the past decade or so. And many topics are from areas still under active development. The goal of the course is to both serve as an introduction to the underlying problems and principles and key methodologies in this field *and* to develop an improved facility for quickly and efficiently navigating the mix of analytical and computational aspects necessary for 'getting up to speed' on approaches in this area.

We will aim to cover roughly a chapter per week or two. Each week, Wednesday and Friday classes will be devoted to lecture, while Monday classes will entail interactive classwork (see below). Corresponding to each chapter will be (i) reading from the textbook, (ii) analytical completion problems, and (iii) data analyses. In addition, there will be a final course project.

1. **Completion Problems:** Problems from the end of each chapter will be assigned as weekend homework. On Mondays, during the interactive class session, students are expected to present the results of their work, which in turn will be used to motivate further discussion. Students are encouraged to work together in teams on these problems. Grading will be based on participation.

2. **Data Analyses:** At a spacing of roughly every few weeks (depending on the particular chapters), a data analysis will be assigned, asking students to implement and/or explore the tools covered during a certain portion of the course. These assignments will involve a somewhat more substantial amount of work than the completion problems, and will be handed in separately and graded.

3. **Course Project:** The course project will ask you to choose a recently introduced methodology from the literature, one not covered during the course itself, and conduct a thorough investigation of that methodology. Project proposals will need to be approved by the course instructor. Project results will be turned in in the form of a written report, including a summary of background and the proposed methodology, implementation details, and description of all simulations, analyses, etc. Further details will be made available by mid-semester.

4. **Grading:** The final grade for the course will be determined according to the following formula.

$$\text{Final Grade} = 0.30\times \text{ Analytical Problems } +0.40\times \text{ Data Analyses } +0.30\times \text{ Final Project}$$

Please Note: Students are responsible for knowing, and abiding by, the provisions of the GRS Academic Conduct Code, which is posted at

http://www.bu.edu/grs/academics/resources/adp.html.

Violations of the code are punishable by sanctions including expulsion from the University.

Course Syllabus (tentative):

1. Week 1: Chapter 2 / Chapter 3
2. Week 2: Chapter 3 cont. (No class Monday)
3. Week 3: Chapter 4
4. Week 4: Chapter 5
5. Week 5: Chapter 7
6. Week 6: Chapter 8 (No Class Monday; Tuesday=Monday)
7. Week 7: Chapter 9
8. Week 8: Chapter 10
9. Week 9: Chapter 15 / Chapter 16
10. Week 10: Chapter 11
11. Week 11: Chapter 12
12. Week 12: Chapter 13
13. Week 13: Chapter 14
14. Week 14: Chapter 17 (No Class Monday; Thursday=Monday)
15. Week 15: Wrap-up on Mon/Wed