BIOSTATISTICS 830 (Special Topic)

MODEL SELECTION

Fall 2008 (3 Credits)

Instructor:
Bin Nan
M4055 SPH II
Phone: 763-5538
E-mail: bnan@umich.edu
Class web: http://www.sph.umich.edu/∼bnan/bio830/

Lectures:
Tue & Thur 10:30am - 12:00pm. Room: M1138 SPH II

Office hours:
Tue & Thur 1:00 - 2:30 pm, or by appointment

Textbooks:
There is no required textbook, but the following books may contain some relevant materials:

- *The Elements of Statistical Learning*, by T. Hastie, R. Tibshirani, and J. Friedman, Springer-Verlag, New York (2001).
- *Model Selection and Multimodel Inference*, by K. P. Burnham and D. R. Anderson, Springer, New York (2002).
- *Concentration Inequalities and Model Selection*, by P. Massart, Ecole d'Ete de Probabilites de Saint-Flour XXXIII-2003. Springer, Berlin (2007).

Prerequisites:
Stat 611, or Stat 621, or permission of instructor

Course Material:
Model selection is a classical area in statistics in which the number of regression coefficients p is often assumed to be fixed while the number of observations n can be large. In the past decade, however, much attention has been drawn to the situation where $p \gg n$, the so-called large p, small n problem that is particularly relevant to a few biostatistical research areas such as genomic studies and brain imaging. There is an explosive literature on the statistical methodologies that deal with a variety of large p, small n problems, but understanding the fundamental statistical properties has been challenging, and many still remain unknown. Rather than providing an exhaustive survey of

the methodology, this special topic course will mainly focus on the most recent theoretical developments (both asymptotic and finite-sample) in this area to help students understand the statistical properties of a few widely applied methods, for example, the Lasso method, and the tools of deriving such statistical properties. The course material will largely be taken from the recent journal publications. Two major aspects of model/variable selection, the regularization and the large-scale multiple testing, will be discussed. Related algorithms will also be discussed.

The instructor will lecture about 85% of the course. The remaining time will be devoted to the discussions on topics presented by students.

Homework:
About 3$\sim$5 assignments.

Exams:
No written exams!

Project:
Depending on the class size, it could be a group project. The final product consists of a paper about 10 pages based on literature review and a talk toward the end of the term. A list of potential topics and papers will be provided.

Grading:
Attendance: 20%; Homework: 20%; Project: 60%

Academic Integrity:

The faculty of the School of Public Health believes that the conduct of a student registered or taking courses in the School should be consistent with that of a professional person. Courtesy, honesty and respect should be shown by students toward faculty members, guest lecturers, administrative support staff and fellow students. Similarly, students should expect faculty to treat them fairly, showing respect for their ideas and opinions and striving to help them achieve maximum benefits from their experience in the School.

Student academic misconduct refers to behavior that may include plagiarism, cheating, fabrication, falsification of records or official documents, intentional misuse of equipment or materials (including library materials), and aiding and abetting the perpetration of such acts. The preparation of reports, papers, and examinations, assigned on an individual basis, must represent each student's own effort. Reference sources should be indicated clearly. The use of assistance from other students or aids of any kind during a written examination, except when the use of aids such as electronic devices, books or notes has been approved by an instructor, is a violation of the standard of academic conduct.