August 20, 2007

Syllabus for STAT 8410 Fall 2007

STATISTICAL THEORY OF BIOINFORMATICS

Instructor: Sounak Chakraborty
Class time: 1.00 pm - 1.50 pm M W F
Location: MDLBH 10
Office: 209F Middlebush Building

Hours Available Outside of Class: Friday 2.00 am - 3.00 pm. Outside office hours by appointment only.
E-mail: chakrabortys@missouri.edu

TA: TBD
Help Session from TA: TBD
E-mail: TBD

Course Web-Page: www.stat.missouri.edu/~chakrabortys

Course Content:

- Review of some elementary statistics.
- Estimation and testing of hypothesis.
- Poisson processes, Markov chains, random walks.
- Analysis of one and multiple DNA sequences.
- BLAST.
- Hidden Markov models and application in protein families.
- Evolutionary models.
- Phylogenetic tree estimation.
- Gene expression, microarrays, multiple testing (if time permits).


Some Valuable References:

Bioinformatics and Computational Biology Solutions Using R and Bioconductor (Statistics for Biology and Health) by Robert Gentleman, Vincent Carey, Wolfgang Huber, and Rafael Irizarry (Hardcover - Aug 31, 2005)

Bioinformatics and Computational Biology Solutions Using R and Bioconductor (Statistics for Biology and Health) by Robert Gentleman, Vincent Carey, Wolfgang Huber, and Rafael Irizarry (Hardcover - Aug 31, 2005)

Statistical Analysis of Gene Expression Microarray Data by Terry Speed

Bayesian Inference for Gene Expression and Proteomics by Marina Vannucci, Kim-Anh Do, and Peter Miller (Hardcover - Jul 24, 2006)

Some Important Dates:
Midterm 10/19/07 (Friday) (6.00 p.m. - 8.00 p.m.) MDLB 10.
Final 12/14/07 (Friday) (8.00 a.m. - 10.00 a.m.) MDLB 10.

Please dont ask for any makeup tests or rescheduling the test. If you have any reservation about the date and time of the midterm you should let me know in the first week of the class. I will try to accommodate. Final exam date cannot be changed for any reason.

Prerequisites : STAT 7760 or permission of the instructor.

Grading:

- Homeworks: 30%.
- Midterm: 20%.
- Final: 20%.
- Project and Presentation: 30% of your grade.

Formula for Calculating the Final Average:

\[
\text{Final Average} = \frac{\text{Total score in all HWs \times 30}}{\text{Total HW Points \times 30}} + \frac{\text{Total score midterm \times 20}}{\text{Total midterm Points \times 20}} + \frac{\text{Total score in final \times 20}}{\text{Total final Points \times 20}} + \frac{\text{Total score in Project and Presentation \times 30}}{\text{Total project Points \times 30}}.
\]

Grading:
There will be one homework assignments from every chapter. Each homework will contain 5 to 10 problems. Apart from that there will be one midterm exam, and one (not cumulative) final exam. Both midterm and the final will be fully in class.

Project and presentation will be a very important part of this course. Since the field of Bioinformatics is changing in a rapid rate I believe to remain competitive in this field some literature study and hand on experience in working with real life data is necessary. You need to form a group of 2 students to work on a project. After you form the group I will assign a project to each groups. I will give you initial references and study materials, but I would strongly encourage to browse through current journals. I will setup a one to one meeting with each group and give some initial pointers for each project. After that you are encouraged to work on your own but if you stumble on any difficulty you are most welcome to consult with me. Since every project will involve some degree of computing I
would suggest you chose a balanced group. Computing can be done with R, Matlab, C, Fortran, according to your choice. I am proficient in all these 4 so I can easily evaluate it. Project report should be within 30 pages, including tables, figures and references. I will give you a LaTeX template and would suggest strongly to use it. Those who are not proficient in latex can write it in MSWORD, (... but still LaTeX is my favorite...). **Your project write up is due on 11/30/2007.** Submit a print out and also email me the soft copy and computer codes. No late submission will be accepted. Each presentation should be of length 30 min. Your project report will carry 15 points and presentation will carry 15 points. Your presentation will be evaluated by me and by your classmates also. **Do not plagiarize or copy from any source like, internet, book, journal etc in your report. Every report I will filter using anti plagiarism software any indication of plagiarism will be punished with an E grade.**

I will request University of Missouri Bioinformatics Consortium to give each group an access to our Lewis and Clark super computer for computing. If they approve then you can use it for your computational need.

I will be quite generous with grades, provided you show me your sincerity and hard work. Based on the performance of the class I may curve but I make no promise that I will do so. The following is a chart of possible final averages and the LOWEST POSSIBLE grade that final average would correspond to. In general

<table>
<thead>
<tr>
<th>Final Average</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>90 - 100</td>
<td>A</td>
</tr>
<tr>
<td>86-89</td>
<td>B+</td>
</tr>
<tr>
<td>80-85</td>
<td>B</td>
</tr>
<tr>
<td>76-79</td>
<td>C+</td>
</tr>
<tr>
<td>70-75</td>
<td>C</td>
</tr>
<tr>
<td>60-69</td>
<td>D</td>
</tr>
<tr>
<td>- 59</td>
<td>E</td>
</tr>
</tbody>
</table>

**Few notes on homework assignments, exams, and projects**

(1) You are not encouraged to work together on homework assignments and any kind of sharing or copying homeworks will be dealt with utmost severity.

(2) Partial credit will be given to partially correct answers, so it is critical that you show all of your work on exams and assignments turned in.

(3) Homework must be neat! If the grader has to spend any time trying to understand what is written, no credit will be given for the problem.

(4) Homework solutions will be posted after 8.00 pm the day it is due on my webpage. Beside this I am also going to post the classnotes, syllabus, Data, and other course related materials in my course webpage. Please visit them regularly to obtain the necessary materials.

**Course Policies**
(1) Exams and homework assignments MUST be taken/turned in at the begin-
ning of the class on the dates given above. Late homework will not be accepted.
Students will only be excused from taking an exam on the date it is given for a serious reason
(poor health, death in the family, etc.) that is verifiable. For this reason, it is absolutely
critical that you not make travel plans which conflict with any of the dates above. Travel
is NOT a serious reason to miss an exam.

2) If you feel that an exam/homework assignment was unfairly graded, you have one week
from the time the exam is returned to make this clear to me.

3) Collecting graded HWs and exams is your responsibility not mine. If you dont collect
your HW and exam in timely fashion and later claim that you have submitted them but
was not graded and you cannot find the old HWs in the pile I cannot award you any credit
for it.

4) It is a part of my TAs job to assist you with your HWs and other questions you may
have. If the TA is not at the office at the designated help session or not answering your
emails let me know. Keep in mind not all doubts can be resolved through mail, so I would
strongly recommend to meet with the TA personally to sort your doubts. Also remember
the TA is a student just like you who has his/her own classes, research, and studies, so dont
expect them to entertain you, at every outside office hours appointment you ask.

All course related materials are available online from my web-
site, www.stat.missouri.edu/~chakrabortys. Dont create or
be in any confusion about the dates, HWs, Soln, syllabus. You
can download every bit of material from the course website.
It is your responsibility to check the course website regularly
for updates.