

# Statistical Methods in Bioinformatics (01:960:691:01), Spring 2003

**Instructor:** Rebecka Jornsten, [rebecka@stat.rutgers.edu](mailto:rebecka@stat.rutgers.edu)

Class hours: MW 9.50-11.10, Serin 385E - note room change for Monday

Office hours: 451 Hill Center, Wed 2.30-3.30

Class homepage: <http://www.stat.rutgers.edu/~rebecka/Stat691>

- You will be graded on HWs & lab reports (30%), a take-home exam & class participation (30%), and presentation of a final project (40%).
- Given the diverse background of the students in this class, you will be expected to focus on different areas of the topics covered.
- Companion papers will be discussed in some lectures.
- Labs in the R-language will be posted and handed out during the semester. These labs will require minimum amount of coding. Instead I want you to focus on results, comparison of methods and clear presentation in lab reports.
- Project: A project can consist of e.g. critical review of a series of papers on a specific topic, or implementation and comparison of methods on real and simulated data sets. Examples of topics: Boosting, Bagging, Current state of classification methods in gene expression data, Methods for Dimension reduction, ...
- Project ctd: You will be required to write a report and present a poster (on the date of the scheduled final exam). Make sure your presentation caters to both the “layman” and the “expert”. Your report will be posted on the class homepage for the benefit of the other students in the class.
- We will cover many topics in this course, but far from all important statistical methods used in bioinformatics research (e.g. sequence alignment, gene finding).

## **Text:**

Required: Friedman, Hastie, Tibshirani. *The Elements of Statistical Learning*

Additional material: lecture notes, handouts.

Suggested: M. Waterman, *Introduction to Computational Biology*, P. Bardi, *Bioinformatics*, J.A. Rice, *Mathematical Statistics and Data Analysis*, Rubin et al, *Biological Sequence Analysis*, Venables and Ripley, *Modern Applied Statistics with S-plus*

### Preliminary Course Outline.

Subject to change!

With such a diverse group of students it's hard to predict the pace of the course.

Lecture notes and other additional material will be posted for each lecture.

Week/Date	Topics	Reading	NB
Jan 22	Introduction	Handout	
Jan 27	Review Prob	Handout	
Jan 29	Review Stats	Handout	
Feb 3	Array techniques	Handout	
Feb 5	Classification - NN, LDA	2.1-2.5, 4.1-4.3, 13.3	R-tutorial
Feb 10	Logistic regression, model selection	4.4, 7.1-7.4	Lab 1
Feb 12	More on discriminant analysis, CART	4.3, 12.6, 12.7, 9.2	
Feb 17	Selection criteria	7.7, 7.8	
Feb 19	TBA		
Feb 24	Bootstrap, validation	7.10	Lab 2
Feb 26	More on validation, Discussion		
March 3	SVM	12.1-12.2	
March 5	SVM	12.3	
March 10	Bagging	8.7	
March 12	More on Bagging, Bayesian model averaging	8.6, 8.8	Lab 3
March 17,19	Spring Break		
March 24	Boosting	10.1-10.4	
March 26	Boosting	10.5-10.7	
March 31	Multiple testing	Handout	
April 2	Multiple testing	Handout	Lab 4
April 7	Discussion		
April 9	Clustering - k-means, VQ, PAM	14.3	
April 14	Trees, SOM	14.3-14.4	
April 16	Validation indices	Handout	
April 21	More on validation	Handout	Lab 5
April 23	Graphical models - Networks	Handout	
April 28	More on Networks	Handout	
April 30	Wrap-up		