



HST.480: BIOINFORMATICS AND PROTEOMICS: AN ENGINEERING-BASED PROBLEM SOLVING APPROACH

Abstract:

Interdisciplinary and hands-on approach the topics of bioinformatics and proteomics. Lectures and labs cover sequence analysis, microarray expression analysis, Bayesian methods, control theory, scale-free networks, and biotechnology applications. Designed for those with a computational or engineering background, it will include current real-world examples, actual implementations, and engineering design issues. Where applicable, engineering issues from signal processing, network theory, machine learning, robotics and other domains. New research areas explored using current literature and book chapter materials written by instructors. Guest lecturers from industry and academia. Graduate students particularly encouraged to register. Graduate level problem sets available for those who want them.

Date: Winter 2005

Credits/Hours:

Four weeks: TR, 11:00 am-12:30 pm

Total hours: 12 scheduled + estimated 28 outside = 40

Weekly load: 3 scheduled + estimated 7 outside = 10

Units: 3 credits (1-0-2) , Level U

Audience: upper undergraduate/first year graduate student

Student Prerequisites:

7.012 or equivalent recommended

6.003 or equivalent recommended

6.041 or equivalent recommended

Grades:

P/D/F.

Course directors: Gil Alterovitz, Marco Ramoni, Manolis Kellis (and other instructors- see syllabus)

Time/Location:

Lecture TR, 11:00a-12:30p, Room 3-133.

Lab (optional), 9:00a-10:40a, Room 37-312.

Syllabus

Date	Time	Location	Theme	Topic	Instructor	Readings	Assignments
Tue, January 4, 2005	11:00am- 11:45pm	3-133	Introduction	Course Introduction, Review of Modern Biology I	Gil Alterovitz	<i>Focus on Robotics and Intelligent Systems Research</i> book chapter [1], sections I-II.	
Tue, January 4, 2005	11:45am- 12:30pm	3-133	Abstraction Level 1: Sequence	Introduction to Bioinformatics Laboratory / Bioinformatics in the Computer Industry	Rob Henson / Gil Alterovitz	MIT guide [2]. Bioinformatics articles [3, 4].	
Thurs, January 6, 2005	11:00am- 11:45pm	3-133	Abstraction Level 1: Sequence	Review of Modern Biology II	Gil Alterovitz	<i>Focus on Robotics and Intelligent Systems Research</i> book chapter [1], sections I-II.	Lab 1 handed out
Thurs, January 6, 2005	11:45am- 12:30pm	3-133	Abstraction Level 1: Sequence	Sequence Analysis: Motif and Regulation	Manolis Kellis	<i>Nature</i> articles [5] and <i>Journal of Computational Biology</i> paper [6].	
Tue, January 11, 2005	11:00am- 11:45pm	3-133	Abstraction Level 1: Sequence	Sequence Analysis: Genes and Genome	Manolis Kellis	<i>Nature</i> article [7, 8]	
Tue, January 11, 2005	11:45am- 12:30pm	3-133	Abstraction Level 1: Sequence	Sequence Analysis: Gene Evolution	Manolis Kellis and Robert Berwick	<i>Nature</i> article [7]	

Date	Time	Location	Theme	Topic	Instructor	Readings	Assignments
Thurs, January 13, 2005	11:00am- 11:45pm	3-133	Abstraction Level 2: Expression	Microarray Expression Data Analysis	Marco Ramoni	<i>Machine Learning</i> paper [9] and <i>Statistical Science</i> article [9].	Lab 1 due, Lab 2 handed out
Thurs, January 13, 2005	11:45am- 12:30pm	3-133	Abstraction Level 2: Expression	Machine Learning: Bayesian Methodologies	Marco Ramoni	<i>PNAS</i> article [10]	
Tue, January 18, 2005	11:00am- 12:00pm	3-133	Abstraction Level 2: Expression	Bioinformatics in the Biotech Industry	Nanguneri Nirmala	<i>Nature Reviews Drug Discovery</i> paper [11]. <i>Science</i> article [12].	
Tue, January 18, 2005	12:00pm- 12:30pm	3-133	Abstraction Level 4: Systems/Misc	Control and Feedback in Systems	Gil Alterovitz	<i>Bioinformatics</i> paper [13]	
Thurs, January 20, 2005	11:00am- 11:45pm	3-133	Abstraction Level 4: Systems/Misc	Scale-free Networks I	Paola Sebastiani	<i>Science</i> [14]and <i>Nature</i> [15] articles.	Lab 2 due, Lab 3 handed out
Thurs, January 20, 2005	11:45am- 12:30pm	3-133	Abstraction Level 4: Systems/Misc	Scale-free Networks II	Paola Sebastiani	<i>PNAS</i> article [14]. <i>Bioinformatics</i> paper [16].	
Tue, January 25, 2005	11:00am- 11:45pm	3-133	Abstraction Level 3: Proteomics	Statistical Models and Stochastic Processes in Proteomics	Gil Alterovitz	<i>Focus on Robotics and Intelligent Systems Research</i> book chapter [1], sections IV-V.	

Date	Time	Location	Theme	Topic	Instructor	Readings	Assignments
Tue, January 25, 2005	11:45am- 12:30pm	3-133	Abstraction Level 3: Proteomics	Signal Processing for Proteomics	Gil Alterovitz	<i>Proteomics</i> article [17]	
Thurs, January 27, 2005	11:00am- 12:00pm	3-133	Abstraction Level 3: Proteomics	Biological Methods, Automation, Robotics	Gil Alterovitz	<i>Focus on Robotics and Intelligent Systems Research</i> book chapter [1], sections III.	Final project due, Lab 3 due (Fri, 5p)
Thurs, January 27, 2005	12:00pm- 12:30pm	3-133	Conclusion	Project Discussion and Wrap-up	Gil Alterovitz		

References

1. Alterovitz, G., E. Afkhami, and M. Ramoni, *Robotics, Automation, and Statistical Learning for Proteomics*, in *Focus on Robotics and Intelligent Systems Research*, F. Columbus, Editor. 2005 (In press), Nova Science Publishers, Inc.: New York.
2. Moormans, M.W., *Matlab on Athena, Technical Publication AC-71*, <http://web.mit.edu/olh/Matlab/TOC.html>. 2001, Massachusetts Institute of Technology: Cambridge, MA.
3. Pochet, N., et al., *Systematic benchmarking of microarray data classification: assessing the role of nonlinearity and dimensionality reduction*. *Bioinformatics*, 2004.
4. Kasturi, J., R. Acharya, and M. Ramanathan, *An information theoretic approach for analyzing temporal patterns of gene expression*. *Bioinformatics*, 2003. **19**(4): p. 449-58.
5. Kellis, M., et al., *Sequencing and comparison of yeast species to identify genes and regulatory elements*. *Nature*, 2003. **423**(6937): p. 241-54.
6. Kellis, M., et al., *Methods in comparative genomics: genome correspondence, gene identification and regulatory motif discovery*. *J Comput Biol*, 2004. **11**(2-3): p. 319-55.
7. Kellis, M., B.W. Birren, and E.S. Lander, *Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae**. *Nature*, 2004. **428**(6983): p. 617-24.
8. Jaillon, O., et al., *Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype*. *Nature*, 2004. **431**(7011): p. 946-57.
9. Sebastiani, P., I.S. Kohane, and M.F. Ramoni, *The role of machine learning in the post-genomic era*. *Mach Learning*, 2003. **52**(1-2): p. 5-9.
10. Ramoni, M.F., P. Sebastiani, and I.S. Kohane, *Cluster analysis of gene expression dynamics*. *Proc Natl Acad Sci U S A*, 2002. **99**(14): p. 9121-6.
11. Kramer, R. and D. Cohen, *Functional genomics to new drug targets*. *Nat Rev Drug Discov*, 2004. **3**(11): p. 965-72.
12. Lawler, A., *Diabetes research. Broad-Novartis venture promises a no-strings, public gene database*. *Science*, 2004. **306**(5697): p. 795.
13. Rangel, C., et al., *Modeling T-cell activation using gene expression profiling and state-space models*. *Bioinformatics*, 2004. **20**(9): p. 1361-72.
14. Goh, K.I., et al., *Classification of scale-free networks*. *Proc Natl Acad Sci U S A*, 2002. **99**(20): p. 12583-8.
15. Bilke, S. and C. Peterson, *Topological properties of citation and metabolic networks*. *Phys Rev E Stat Nonlin Soft Matter Phys*, 2001. **64**(3 Pt 2): p. 036106.
16. Rzhetsky, A. and S.M. Gomez, *Birth of scale-free molecular networks and the number of distinct DNA and protein domains per genome*. *Bioinformatics*, 2001. **17**(10): p. 988-96.

17. Baggerly, K.A., J.S. Morris, and K.R. Coombes, *Reproducibility of SELDI-TOF protein patterns in serum: comparing datasets from different experiments*. *Bioinformatics*, 2004. **20**(5): p. 777-85.

List of Speakers (Alphabetical)

Name	Affiliation
Gil Alterovitz	HST Medical Engineering Medical Physics-Electrical Engineering and Computer, Graduate Student/Whitaker Fellow
Robert Berwick	Professor, Electrical Engineering and Computer Science, MIT
Rob Henson	Director of Bioinformatics Group, Mathworks (Matlab).
Manolis Kellis	Assistant Professor, Electrical Engineering and Computer Science, MIT Other affiliations: Computer Science and Artificial Intelligence Laboratory (CSAIL), and the Broad Institute of MIT and Harvard.
Nanguneri Nirmala	Functional Genomics Group, Novartis Institutes for BioMedical Research
Marco Ramoni	Assistant Professor of Pediatrics and Medicine, Harvard Medical School Other affiliations: Assistant Professor of Oral Medicine, Harvard School of Dental Medicine Associate Director of Bioinformatics, Harvard Partners Genome Center Children's Hospital Informatics Program
Paola Sebastiani	Associate Professor, Department of Biostatistics, Boston University School of Public Health

Labs

Lab 1	Sequence Analysis
Lab 2	Expression theme based Lab
Lab 3	Proteomics and Systems theme-based Lab
Final Project	Student selected based on one of the four themes of course

Grading

Labs	40%
Final Project	50%
Participation	10%