ECE 09.402

SPECIAL TOPICS: INTRODUCTION TO BIOINFORMATICS
(3 SCH - UNDERGRADUATE)

ECE 09.504
SPECIAL TOPICS: BIOINFORMATICS
(3 SCH - GRADUATE)

FALL 2011
There is no shortage of definition for bioinformatics. Loosely speaking, bioinformatics is the field of applying computational techniques, from mathematics, statistics, and machine learning, to the vast amounts biological – but most specifically genomic – data. While some people refer to bioinformatics only in the context of collection, storage, organization and access of such biological data within large databases, our view of bioinformatics will include – in fact focus on – development of algorithms and models to analyze such vast quantities of data to make some sense of it.

Bioinformatics is one of the fastest growing fields of science, in part due to the massive amount of data generated by the availability of fast next generation machines that are capable of quickly sequencing complete genomes of species in a relatively short period of time. Just as an example, the human genome, so called the code of life, consists of about 3 billion letters (base pairs, nucleotides) that encode about 23,000 genes, which are responsible for coding proteins which in turn are responsible for everything from your eye color to curl of your hair to your risk of developing a heart disease or Alzheimer’s. Humans are just one of 1.3 million identified of unknown number of species on this planet, of which a miniscule 1000 of them have been sequenced. New species are being discovered and sequenced as we speak, adding to the vast amount of data we already have. Bioinformatics, therefore, is the study of storing, organizing and analyzing this huge data.
Bioinformatics can hardly be overstated – it allows us to determine the role and function of the genes and the proteins they encode, which in turn can help us better understand the mechanism of life, and allow us to develop tools to fight diseases that are caused by (mutations of) certain genes.

This introductory class will introduce you to the basic concepts of bioinformatics, including the various databases and computational tools available to analyze those databases. However, we will also spend time on developing such tools, many of which are primarily within the domain of machine learning / statistical learning. This class has four main goals:

1. To provide you with basic biological background of genomics, so that you can fully understand the necessity of developing computational tools.
2. To introduce you to commonly used bioinformatics databases (such as NCBI’s GenBank, PDB) and the computational tools (such as BLAST) used to analyze genomic data from such databases.
3. To equip you with mathematical and statistical (machine learning) techniques commonly used in bioinformatics, which will not only help you understand the underlying principles of the popular computational tools, but also allow you modify them or even develop new tools to solve more challenging problems in bioinformatics.
4. Inform you regarding the latest developments in this increasingly important and growing area, and sensitize you to contemporary issues on biotechnology and bioinformatics, including the ownership and use of genetic data.

Upon successful completion of this class, you will be able to

1. Access bioinformatics databases, extract various genomic data from such databases, determine and use appropriate tools, and analyze and interpret such data;
2. Gain a working knowledge of fundamental topics in machine learning, such as k-nearest neighbors, k-means, support vector machines, and hidden Markov models, which will allow you to implement and modify existing algorithms and develop new algorithms for various applications of bioinformatics.
3. Explain and present the relevant concepts of bioinformatics, the tools used and developed for genetic data analysis as well as the interpretation of the outcome of such analysis using proper technical language, both in oral and written formats and venues.
4. Discuss various recent developments in this rapidly growing area, such as gene therapy and synthetic life, and form your opinion regarding ethical issues surrounding this topic, including private ownership of genetic data.

COURSE PREREQUISITES

- Basic knowledge of probability, statistics and random variables, linear algebra – See introductory material
- Calculus III or Math for Engineering Analysis,
- Signals and systems / digital signal processing
- Expertise of MATLAB or C/C++
- Enthusiasm, genuine interest, and willingness to put forward extra effort
- Time, patience, perseverance

INTRODUCTORY / BACKGROUND MATERIAL

Many of you might have seen the mathematical background material sometime ago, or perhaps had very little exposure to these areas. While we do not have time review these topics, I have provided extensive material on the class webpage. These include my own notes, links to some very good video lectures, as well as other tutorials. You are encouraged to review this material throughout the semester as we introduce more advanced topics.

CLASS MECHANICS

This class will meet twice a week, for 75-90 (undergraduate-graduate) minutes each. Any class period that cannot meet (due to holidays, weather, etc.) will be made up. Due to multidisciplinary nature of this class, we will have several guest speakers and lecturers, as well as some intriguing and thought provoking video features. We will have homework, research and implementation exercises, oral reviews, one midterm and/or final exam and a final project. We will also review one of the most fascinating races of recent science history: sequencing the human genome. You will be asked to read two books, written from diametrically opposing views of this race.
**Attendance Policy & Estimated Amount of Work**

Attendance is absolutely necessary for success in this class, and therefore is required. Since this is a senior/graduate level class, I will not take regular attendance, but only occasional random ones, which will be considered in assigning final grades. If you are absent on the day you are randomly called for oral review (or for roll call), you will get a zero for that activity – unless you have an excused absence or an extenuating circumstance. I will not call on someone for an oral review if that person informed me of his/her excused absence ahead of time. A missed exam may not be retaken with the exception of most serious and extenuating circumstances that require official and written proof of such circumstances.

**Excused Absence** is one where you have given me at least 48 hours of written notice (e-mail is acceptable) of your absence. You may have one – and only one – excused absence during the semester, though it can be for any reason.

**Extenuating Circumstances** are those that are truly beyond your control, such as sudden illness, or death of family member. Written documentation must be provided for an extenuating circumstance to be valid (such as a letter from a physician, or an obituary/funeral house notice). Undocumented cases will not be honored.

**Tardiness:** You will be considered present if you are in class during the first 10 minutes of the class, and remain in class during the entire (remaining) duration of the class. A quiz or oral review missed due to tardiness will be counted towards your excused absence. Any additional absences / tardiness will result in a zero grade for the missed activity.

Note that this class is a biologically, mathematically and computationally intensive graduate level course. You will be learning a substantial amount of cutting edge material, and you will be writing simulation programs to test them. Expertise (not just familiarity) with Matlab or some other programming language is absolutely essential. Because this is an upper level / graduate course, you will also be expected to do a substantial amount of reading – not only from the text but also from scientific magazines and journals. Successful completion of this course will demand significant amount of time commitment from you, a good portion of which may be spent on reading and algorithm implementation. As a rule of thumb, expect to spend three - four hours for each hour we spend in class, i.e. 9 – 12 hours a week on top of class meetings.

If you miss a class, you are responsible for any missed material, and given the pace and level of this course, even a single missed lecture will be difficult to catch up. **So don’t miss class!**

**Reason for Taking This Class**

Bioinformatics is a challenging, yet intellectually rewarding course! I therefore assume that you are in this class because you are truly interested in the subject. You should not be in this course if you signed up for it for any other reason than enthusiastic and genuine interest in learning the contents of this course.

**Team Policy for Class Related Work**

You are not only allowed, but in fact encouraged to work in teams (usually of no more than two) for most class related work, including certain homework assignments (unless stated otherwise) – but not for exams / quizzes. Whether you can use a team for the final project will depend on the complexity of the project. You are free to form / deform as many teams as you wish during the semester for all homework / implementation assignments. As long as all team members contribute equally and their names appear on the homework assignments, one can be submitted by each team. Each team member, however, needs to explicitly state the sections that were his/her primary responsibility. Team members may inform me – under the condition of anonymity – of other team members who are not equally cooperating or participating in team effort.

**Homework / Research & Implementation Assignments**

There will be irregular homework / research / implementation assignments that will challenge you, however, you will realize that you learn a lot from these exercises. As an added bonus, you will notice that your analytical thinking, problem solving will also improve significantly, not to mention your math skills. Assignments must be neatly and professionally prepared and **typed**, using proper and formal technical writing language. See IEEE style guide for guidance. All assignments will be due one or two weeks from the day they are assigned, unless indicated otherwise. **Late Policy:** Late submissions will not be accepted.
**Reading and Video Assignments**

As in most graduate level courses, you will be asked to read certain portions of the course material on your own from your text, other texts, or research articles. I will also ask you to watch certain video lectures that are available on the web (http://videolectures.net). Finally, there are two popular science books that I want you to read throughout the semester. These books are related to a race for finishing the sequencing of the human genome, written from two different perspectives. We will then discuss the implications and the ethical issues that surround them towards the end of the class. Homework, exam and oral review questions will be drawn from such assignments.

**Course Project**

A final project to help you put all course-developed skills to work will be assigned. You may choose from one of the three options below, on which you will have a minimum of one month to work. All project ideas must be pre-approved by me for appropriate scope and depth.

**Graduate Students: (Undergraduate students get 15% bonus for selecting this option)**

Develop a new technique, either from scratch, or by suitably modifying an existing technique for a specific problem of your interest; test it on an appropriate benchmark database.

**Undergraduate Students:**

Identify a new bioinformatics / machine learning algorithm not covered in class – relevant to class material – from a recent journal article (>2006). Implement the algorithm and evaluate it on an appropriate database.

**All students:**

Suggest your own project topic. Must be pre-approved by the instructor.

Furthermore, a paper submission (possibly after semester ends) to a conference or journal is required for all graduate students to get an A/A- from this course. Many related conferences may have deadlines after the semester ends. Therefore, a letter of intent to submit, and a draft version of the paper – in the appropriate conference required format - that shows justifiable progress on the final project that would warrant a reasonable chance of acceptance will be sufficient for this purpose. Acceptance of your paper to the conference is not a condition for the “A,” just the submission. If your paper is accepted, financial assistance will be available to cover – at least a part of – the travel and attendance costs, provided by the instructor and/or the department. Here are some related conferences and their submission deadlines:


**Note that without a paper submission, the best attainable graduate grade is a B+.**

**Class Ethics:**

- No eating /drinking in class (except bottled water). Absolutely no sandwiches, pizza, hoagies, soda, etc.! Please time yourself accordingly.

- **Cell phones must be kept outside of class or shut-off during class. No exceptions!** If your cell-phone rings during class / lab (or you use it in any other way, such as texting), you will be asked to leave and counted as unexcused absent. It will also cause very difficult-to-repair damage to “professionalism” part of your grade (see below). Furthermore, I reserve the right to answer any cell phone going off in the class.

- No web surfing, instant messaging and / or other unrelated use of computers, when we use computers in class / labs.

- In-class discussions are welcome, and in fact encouraged, within the limits of mutual respect and courtesy. I also expect proper business conduct in the class – I am all about in-class discussions, so long as such discussions include the entire class, and not between two students.

- You are responsible for checking the class web page often for announcements, homework / exam solutions.

- You are encouraged to work with other students for all exercises, except exams and quizzes.

- Please come to class in proper (business casual) attire. Slacks and shirts are fine; PJs and flip-flops (yes, it did happen) are not!
• Although I do not anticipate, and certainly hope that it will never be an issue, it is my responsibility to re-
mind you that academic dishonesty – in any form, shape or manner – will not be tolerated, and will be dealt
with according to university rules and regulations. In general, presenting any work, or a portion thereof,
that does not belong to you, as if it does – or even attempting to do so – is considered academic dishonesty.
**DO NOT DO IT, DO NOT EVEN ATTEMPT IT!** See the pledge of honesty at the end of this syllabus.

**E-MAIL ETIQUETTE & POLICY:**

In general, I prefer that class related questions be asked in the class, so that everyone can benefit from the dis-
cussion. If your answer requires a longer time to answer, then please ask in person. Use e-mail as a last resort only.

If you do use e-mail, I expect you to follow proper business etiquette for electronic communications, including
a formal greeting (for example, “Dear Dr. Polikar”, and not “Hey!” or “Listen, Robi, help me out here...”), formal
language and formal signature line (including your name, last name and Banner ID). E-mails that do not follow
proper etiquette will not be answered, and will affect the professionalism portion of your grade.

Also, and this in very important: make sure that your Rowan e-mail account is active. I will use your Rowan e-
mail address exclusively, and cannot / will not follow up with messages that bounce back for over quota issues or
non-Rowan addresses. All e-mails sent to me **MUST come from your Rowan e-mail address**, preferably with a
subject line starting with “Bioinformatics”. My e-mail client is configured to recognize all e-mails coming from
Rowan addresses as legitimate. Any other e-mail address may – and probably will – be automatically classified as
spam and I may not notice it in a timely manner.

**GRADING SCALE**

An absolute grading scheme will be used to assess your final grade:

<table>
<thead>
<tr>
<th>Category</th>
<th>Weight</th>
<th>Grade Range</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exams</td>
<td>35%</td>
<td>100-95: A, 95-90: A-</td>
<td></td>
</tr>
<tr>
<td>Assignments/</td>
<td>25%</td>
<td>89-87: B+, 86-83: B, 82-80: B-</td>
<td></td>
</tr>
<tr>
<td>In-class:</td>
<td></td>
<td>79-77: C+, 76-73: C, 72-70: C-</td>
<td></td>
</tr>
<tr>
<td>Project:</td>
<td>30%</td>
<td>69-67: D+, 66-63: D, 62-60: D-</td>
<td></td>
</tr>
<tr>
<td>Professionalism:</td>
<td>10%</td>
<td>59-0: F</td>
<td></td>
</tr>
</tbody>
</table>

Professionalism includes good academic citizenship, professional conduct, and active class participation.

**OFFICE HOURS & CONTACTING THE PROFESSOR**

I will hold open office hours for this class. This means that you may come in at any time to ask questions if my
office door is open (which it usually is). A couple things to note however: Please do not come in if the door is
closed, even if my light is on, or you know that I am, in fact, inside. My door is typically open, but if it is closed,
that means either I am not in, or I am working on something a

Also, you may always make appointments (for example, if you need my uninterrupted attention for an extended
period of time), or you wish to discuss something in private.

**ACCOMMODATION FOR DISABILITY**

If you have a documented physical and/or learning disability, please feel free to inform me or the Cen-
ter for Academic Success – CAS (director, Ms. Melissa Cox – cox@rowan.edu, or 256-4260) regarding
what kind of accommodation you need to help you succeed in this class. While you are not required to
disclose your disability to me, you must provide appropriate documentation to the CAS to receive official
university assistance. All such requests will be held confidential to the extent possible.

**Course Content**

---

1. Actual examples from previous student e-mails.
2. 10% of project component will come from preparing an appropriately formatted and edited, near final version of the draft conference manuscript that is ready for submission.
The course will consist of three primary parts:

**Fundamentals:**
Definition, structure and functions of DNA / RNA, genes, proteins, transcription and translation, central dogma of molecular biology; major forms of life; primary, secondary, tertiary structure of proteins.

**Processes, Tools & Technologies Used in Bioinformatics:** Bioinformatics databases, polymerase chain reaction, sequencing, local and global sequence alignment, sequence search tools (FASTA, BLAST, etc.) gene and protein expression, microarrays and gene chips, genome annotation, phylogenetics.

**Computational Intelligence for Bioinformatics:** The theory behind computational tools used in bioinformatics: Bayesian analysis, naïve Bayes classifier, clustering methods, hidden Markov models, supervised classification with decision trees, neural networks, and support vector machines.

Please note, however, these topics will not necessarily be discussed in the order they are listed above primarily due to two reasons. First, some of the topics will be presented by guest lecturers, whose schedules are still being determined, and/or may change at any time. Second, some of the computational intelligence techniques will be presented along with the application for which they are indented.

**Instructor Evaluation, Questions, Comments, Suggestions**
Questions, constructive criticisms, comments, and suggestions are always welcome. Please feel free to share your opinions about all aspects of the class: content, math level, workload, instructor’s communication skills (or lack thereof), etc. There will be a box outside of my office for anonymous comments. Feel free to use this box, if you wish to remain anonymous regarding your comments. Also, you may use the “I’ve got something to say” form, available at class homepage for your comments. A copy is attached to this syllabus. I will also give you a mid-semester evaluation form, so that you can have a formal opportunity to voice your concerns or appreciations (if any at all…).
I am having difficulty in understanding the following concepts:

This week’s class was informative / interesting / entertaining / _______ (circle all that apply) because:

This week’s class was confusing / boring / too fast / too slow / _______ (circle all that apply) because:

It would have been much better / beneficial if you could…:

Please continue the following activities as I find them useful in______

While you are at it, please provide your feedback on the following on a scale of 1 – 5,

1: Poor / Strong disagreement with the phrase, 5: Excellent / strongly agree with the phrase

1.  The professor’s ability to communicate in a clear and understandable manner: _____
2.  The professor’s responsiveness to student’s needs, questions and ideas:______
3.  The professor treat students in a professional manner:______
4.  The professor is enthusiastic about the subject and genuinely believes in its importance:____
5.  The professor’s knowledge of the subject material is thorough:_____
6.  The professor is well prepared for the classes:_____ 
7.  The professor’s ability to impart knowledge about the subject is:______
8.  The professor encourages questions and comments during the class session:_____
9.  The professor’s use of the class time is:______
10. The professor actively involves students in the teaching / learning process:_____
11. The professor’s availability outside of class hours is:____
12. The professor satisfactorily answers students’ questions in class and in the office: _____
13. Professor clarifies / repeats material that is difficult to understand:_____
14. Professor makes use of the latest technology to improve student’s learning experience:_____
15. Lecture materials (e.g. slide) are helpful for the understanding of the subject material:____
16. The professor is genuinely concerned that students take valuable experience from the class:____
17. Considering everything, how would you rate this teacher: ______

What do you not like about Dr. Polikar’s teaching, if any, and what would you suggest that he can do improve?

What do you enjoy about Dr. Polikar’s teaching, if any, that he should continue in this and future classes?

“I’ve got something to say!” Course & Instructor Evaluation – Robi Polikar © Fall 2011.
HONESTY STATEMENT

I, _____________________________, have carefully read and understood this document, and clarified any questions or issues I may have (if any) with the instructor. I understand the class expectations, agree to follow the class rules, and in particular I attest that all work I present as my own will be my own. I will not represent other’s work as mine. Collaborations for mutual work notwithstanding, I will not accept nor offer any help from/to anyone for work individually expected of me, and I will not use any unauthorized sources, including but are not limited to books, notes, preprogrammed calculators, the internet, other students, etc. during the exams / quizzes. I understand that showing others’ work as mine, or attempting to do so, is considered academic dishonesty, and such acts are subject to forfeit of any grade obtained from the exam as well as other disciplinary action from the university. I also attest that, I will not share any exam questions with other students - past, current or future - with the understanding that doing so will also be considered academic dishonesty.

___________________________________  __________________
Name                                  Banner ID
___________________________________  __________________
Signature                             Date