BENG 420: Bioinformatics for Engineers  
Fall 2014

Credits 3  
Thursdays 1:30 pm - 4:20 pm  
Robinson Hall, A243

Instructor:  
Qi Wei, PhD  
Assistant Professor  
Department of Bioengineering  
Office: Nguyen Engineering Building, Room 3904  
Email: qwei2@gmu.edu  
Phone: 703-993-5211  
Office hours: Mondays noon-2pm and by appointment

Prerequisite(s):  
Grade C or better in BENG 320 or SYST 320 or ECE 320

Course Description: This course introduces the fundamental techniques and tools for analyzing biomedical data, important for many biomedical engineering problems. Topics include classification, regression, clustering, dimensionality reduction, data representation, pattern matching and algorithm performance evaluation. Students will deepen their understanding of the concepts and gain hands-on experience on data analysis by applying algorithms to analyze and interpret various kinds of standard biological and biomedical data encountered in bioengineering fields. This course will be an innovative course leveraging hybrid learning through a combination of lectures, on-line content, and course project.

Learning Objectives
The following student learning outcomes will be directly assessed:

- Students will have an ability to apply knowledge of mathematics (including differential equations and statistics), science (including biology and physiology), and engineering to solve problems at the interface of engineering and biology.
- Students will have an ability to use computational tools, such as regression analysis, classification algorithms, data reduction methods, pattern recognition algorithms, to analyze and interpret data from living systems.
- Students will have an ability to communicate effectively through writing.

Textbook
There is no textbook for the class. You will be assigned additional reading materials including book chapters and papers throughout the class. Students are required to review online lectures from the Coursera Machine Learning course (https://www.coursera.org/course/ml) according to the schedule indicated below. Please sign up an account on Coursera and register the course in order to view all materials. It is expected that you review the appropriate module before class. Note that Coursera lecture
notes are also available on the website. You should be able to view them after you log in and follow the Course Wiki link on the left.

**Grading (tentative)**

- Course project (25%)
- Homework (50%)
- Participation in discussion (5%)
- Midterm (20%)

**List of Weekly Topics**

<table>
<thead>
<tr>
<th>Wk</th>
<th>Date</th>
<th>Topic</th>
<th>Required reading</th>
<th>Assignment &amp; Project Deadlines</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8/28</td>
<td>Introduction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>9/4</td>
<td>Data Representation and Exploration</td>
<td>Chap 2 and 3</td>
<td>Assignment #1</td>
</tr>
<tr>
<td>3</td>
<td>9/11</td>
<td>Classification I: intro, neural networks I</td>
<td>VIII</td>
<td>Assignment #1 due</td>
</tr>
<tr>
<td>4</td>
<td>9/18</td>
<td>Classification II: Neural networks II</td>
<td>IX</td>
<td>Assignment #2</td>
</tr>
<tr>
<td>5</td>
<td>9/25</td>
<td>Classification: Support Vector Machines</td>
<td>XII, Handout</td>
<td>Assignment #2 due; Project proposal due</td>
</tr>
<tr>
<td>6</td>
<td>10/2</td>
<td>Clustering (k-means)</td>
<td>XIII</td>
<td>Assignment #3;</td>
</tr>
<tr>
<td>7</td>
<td>10/9</td>
<td>Guest Lecture by Dr. Amarda Shehu</td>
<td>II &amp; III</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>10/16</td>
<td>Linear regression with multiple variables</td>
<td>IV</td>
<td>Assignment #3 due; First draft of individual project report due (Introduction and Methods sections)</td>
</tr>
<tr>
<td>9</td>
<td>10/23</td>
<td>Logistic regression (classification vs. fitting)</td>
<td>VI &amp; VII</td>
<td>Midterm; Assignment #4</td>
</tr>
<tr>
<td>10</td>
<td>10/30</td>
<td>Dimensionality reduction, Principal Components Analysis</td>
<td>XIV, handout</td>
<td>Data mining competition</td>
</tr>
<tr>
<td>11</td>
<td>11/6</td>
<td>Independent Components Analysis, Linear Discriminant Analysis</td>
<td>Handout</td>
<td>Assignment #4 due; Assignment #5</td>
</tr>
<tr>
<td>12</td>
<td>11/13</td>
<td>Anomaly detection</td>
<td>XV</td>
<td>Assignment #5 due</td>
</tr>
<tr>
<td>13</td>
<td>11/20</td>
<td>Computational framework of bioinformatics; Project presentation I</td>
<td></td>
<td>Second draft of individual project report due (Results section)</td>
</tr>
<tr>
<td></td>
<td>11/27</td>
<td>Thanksgiving recess. No Class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>12/4</td>
<td>Project presentation II</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>12/10</td>
<td></td>
<td></td>
<td>Final project report due</td>
</tr>
</tbody>
</table>
The instructor reserves the right to make any changes in the course she determines academically advisable. Changes will be announced in class. It is your responsibility to keep up with any changed policies or course schedule.

**Course structure:**
The course will consist of a weekly lecture, homework assignments, an individual project and a midterm exam. The homework assignments will involve programming in Matlab. The midterm exam will be closed book and closed notes. The project will involve hands-on programming of relevant data analysis algorithms using real or simulated data. Project reports that lack substantial interpretation of the results will not receive much credit.

**Attendance:**
Your attendance and engagement are crucial for meeting this course’s objectives. Without a valid excuse deemed justified by the instructor, you are expected to be on time and to attend all sessions. A valid excuse includes, but is not limited to: illness, death in the family, or personal crisis. Notification must be given before class begins. Students are allowed one absence with no questions asked. Students who accrue 2 absences without notifying the instructor will receive a grade reduction.

**Homework:**
There will be assigned homework throughout the semester, which will be posted on Blackboard via myMason. The homework problems may involve some programming and analysis. Homework submitted after the due date will be penalized (15% penalty for each day late). Once homework is discussed in class or the solution is posted, submissions will no longer be accepted either.

**Participation in Discussion:**
Your participation in class discussion will contribute to 5% of the class grade.

**Midterm Exam:**
There will be a midterm exam in the class. The exam will be closed book and notes. They will consist of a mixture of numerical problems as well as essay-type, and multiple-choice type questions. Absence from the exam must be notified ahead of time and alternative arrangements made with the instructor.

**Project:**
The students will be required to do a course project that will integrate the material learnt in the course. The students are encouraged to seek out their own project topics or select from a list of suggested topics. Typically the projects will involve programming and implementation of data analysis algorithms. The student will be expected to seek out and integrate relevant information and demonstrate an ability to apply the methods from the class and evaluate the results. Students will develop sections of the project report throughout the semester and receive feedback, which will be graded.

The project reports should be formatted as a 4(or more)-page, 2-column, IEEE conference paper in 12-point Times New Roman font, with appropriate references in
IEEE format. The project report is expected to consist of four sections: (1) Introduction and Background, (2) Methods and Materials, (3) Results, and (4) Discussion and Summary. Project reports that lack substantial and critical analysis and interpretation of the results will not receive much credit. For the individual project, students are expected to submit drafts of these four sections by the due dates indicated in the schedule. Instructors will provide commentary on the draft, and the revised draft will be due by the dates indicated below. For more detailed description of the expected contents of the project report, please see the project assignment.

**Academic Honesty and Collaboration:**

The integrity of the University community is affected by the individual choices made by each of us. GMU has an Honor Code with clear guidelines regarding academic integrity. Three fundamental and rather simple principles to follow at all times are that: (1) all work submitted be your own; (2) when using the work or ideas of others, including fellow students, give full credit through accurate citations; and (3) if you are uncertain about the ground rules on a particular assignment, ask for clarification. No grade is important enough to justify academic misconduct.

With collaborative work, names of all the participants should appear on the work. Homework problems are designed to be undertaken independently. You may discuss your ideas with others and conference with peers; however, it is not appropriate to give your work to someone else to review. You are responsible for making certain that there is no question that the work you hand in is your own. If only your name appears on an assignment, your professor has the right to expect that you have done the work yourself, fully and independently.

Plagiarism means using the exact words, opinions, or factual information from another person without giving the person credit. Writers give credit through accepted documentation styles, such as parenthetical citation, footnotes, or endnotes. Paraphrased material must also be properly cited. A simple listing of books or articles is not sufficient. Plagiarism is the equivalent of intellectual robbery and cannot be tolerated in the academic setting.

There will be a zero tolerance policy in this course for plagiarism in the written project reports and in homework submissions. Every instance of plagiarism will be reported to the GMU Honor Committee. **No excuses. No exceptions.** If you have any doubts about what constitutes plagiarism, please talk to the instructors.

**Relevant Campus and Academic Resources**

**Disability Services**
If you have a documented learning disabilities or other conditions that may affect academic performance should: 1) make sure this documentation is on file with the Office of Disability Services (SUB I, Rm. 4205; 703 993 2474 ;http://ods.gmu.edu) to determine the accommodations you might need; and 2) talk with the instructors to discuss reasonable accommodations.

**Office of Diversity, Inclusion and Multicultural Education**
SUB I, Rm. 2400; 703 993 2700; http://odime.gmu.edu/

**Writing Center**
Robinson A114; 703 993 1200; writingcenter.gmu.edu.