CSE 5301-001: Data Analysis and Modeling Techniques
Fall 2014

Official Location of This Document:

http://crystal.uta.edu/~zaruba/CSE5301/CSE5301-syllabus.pdf

Short Course Description:

This course introduces concepts and techniques for performing experiments and analyzing their results. Topics cover fundamental statistics, probability and data-representation concepts, inference through hypothesis testing, information theory, queuing models, and selected topics such as capacity planning and bottleneck analysis, clustering and classification, and hidden Markov models with computer science applications as examples.

Prerequisites:

- Graduate standing
- Engineering Probability *(IE 3301) or equivalent undergrad.* statistics/probability theory

Instructor: Gergely Záruba

- Office: 523 ERB
- Phone: (817) 272-3602
- Office hours: Tuesdays and Thursdays 11:00am – 12:00pm
  (other consultations by appointment only.)
- Instructor's e-mail: zaruba@uta.edu
- GTA: TBA

Objectives:

This course briefly covers basic statistics and probability concepts and introduces techniques to model and analyze probabilistic data. This includes basic representation such as Bayesian networks as well as hypothesis testing techniques for data analysis and interpretation. Further, it introduces modeling and analysis techniques for sequential processes, including Markov models, regression analysis, and basic queuing models. All of these techniques will be discussed in the context of common Computer Science problems from a wide range of fields, including Computer Networks, Artificial Intelligence, Machine Learning, Computer Vision, Data Mining, Bioinformatics, etc. In addition, the course will discuss selected advanced topics and applications such as capacity planning and bottleneck analysis, clustering and classification

Outcomes:

Students successfully completing this course will have gained a solid understanding of probabilistic data modeling, interpretation, and analysis and thus have formed an important basis for more advanced courses in computer science as well as for the handling and analysis of data used in real-life applications and research.

Details of Curriculum:

- Class meets Fridays 9am-11:50am (with a ten-minute break) in ERB129
- Class WWW site: http://crystal.uta.edu/~zaruba/CSE5301/
  Note: Please check WWW site for up to date information

- Class mailing list address: CSE5301-ZAR@LISTSERV.UTA.EDU
  Note: students are strongly encouraged to sign up for the mailing list of the class; ***It is your responsibility to verify that you are added to this mailing list. If you have doubts, please send me an email within the first weeks of classes.***
- Text Books:
  There are a wide range of books on this topic, all of which cover many of topics covered in the course and can be used as references for the course. However, none of them covers everything in the course. As a consequence the course does not follow any one specific textbook. However, book recommendations will be made throughout the course.
- Lecture slides will be placed on the web.

Details of Class Policies:

Course Grades:

Tentatively, course grades will be based on the following:

- Homework assignments (50%)
  - There will be five homework assignments spread throughout the course (10% of the grade, each).
  - There will be no make up for assignments/homework. The maximum grade given for assignments/homework will decline by 20% of the total grade each calendar day the assignment/homework is overdue starting razor sharp after the deadline.
  - Homework assignments may contain programming components. The choice of programming language is left to the student. However, in some assignments simulation and data generation components might be provided which will be implemented in C or C++. These components will not be provided in additional languages and thus interfacing with C or C++ (which is possible in most programming languages) might be necessary when a different programming language is used. In all cases, the following limitation will apply to the programming language chosen: All programs must compile and run on university machines (either university servers or the machines in the open OIT laboratories) and instructions regarding how to compile and run the code must be provided with the program submission. In case of doubts regarding the use of a particular programming language or software package, contact the instructor prior to its use.
- Two Midterms: (50% total; 25% each)
  - There will two midterm exams during the semester, on October 10, 2014 and November 21, 2014.
  - There will be no make-up exams!

Pop-Quizzes may be possible throughout the semester; points received (or not received) will apply towards the grade in the "homework" grading part. Tentatively, course grades are determined from the total points (100) earned as follows, (but the instructor reserves the right to "grade over the curve," or even to give everyone the best grade):

90-100: A ; 75-89: B ; 60-74: C ; <60: F

Make-ups:

Make-ups for (non-exam) graded activities may be arranged if your absence is caused by illness or work/personal emergency. A written explanation (including supporting

documentation) must be submitted to your Instructor. If the explanation is acceptable, an alternative to the graded activity will be arranged. Make-up arrangements must be arranged prior to the scheduled due date.

Attendance:

Students are not required but encouraged to attend all class sessions, however they are encouraged as class attendance is strongly correlated to the quality of grade received. No cell phones, loud talking, and sleeping in the classroom, please.

Notes:

- The Instructor reserves the right to modify course policies, the course calendar, and assignment or project point values and due dates.
- All students are expected to be responsible users of the computer systems used for this course.

Accepted file formats for papers/reports:

The Instructor requires the students to turn in their papers and reports either in *.pdf* (Adobe's portable document format – can be generated, e.g., either by *Adobe Distiller* or later versions of *ghostscript*) or in *.ps* (Adobe's Postscript – can be generated, e.g., from Latex source files by *latex* and *dvips* or from the Windows operating systems by installing a virtual postscript printer device and printing the document to a file) formats. Source files (!) must be turned in along with the paper in a zip or a gzip (or tgz, .tar.zip) archive. Students are encouraged to use the Latex language and its appropriate compilers or the Microsoft Office program family (please see the Instructor if you intend to use anything else). If viruses are submitted along with the files a student turns in, the Instructor may degrade the grade of the assignment.

Homework assignments may contain programming components. The choice of programming language is left to the student. However, in some assignments simulation and data generation components might be provided which will be implemented in C or C++. These components will not be provided in additional languages and thus interfacing with C or C++ (which is possible in most programming languages) might be necessary when a different programming language is used. In all cases, the following limitation will apply to the programming language chosen: All programs must compile and run on university machines (either university servers or the machines in the open OIT laboratories) and instructions regarding how to compile and run the code must be provided with the program submission. In case of doubts regarding the use of a particular programming language or software package, contact the instructor prior to its use.

Grievance Procedure:

Anyone feeling that a dispute exists after the grading of any assignment or exam may submit a written grievance. This grievance should identify the item in dispute and arguments supporting the student's position. Grievances must be submitted in writing within two class periods following the return of the assignment. The instructor or GTA agrees to return a written response to the student's grievance within two class periods from receipt of the grievance. If the error is due to wrongful calculation of points, then no grievance needs to be submitted. If a written grievance is received, the instructor and GTA reserve the right to re-grade the entire exam (not just the specific point in question).

Drop policy:

Students may drop or swap (adding and dropping a class concurrently) classes through self-service in MyMav from the beginning of the registration period through the late registration period. After the late registration period, students must see their academic advisor to drop a class or withdraw. Undeclared students must see an advisor in the University Advising Center. Drops can continue through a point two-thirds of the way through the term or session. It is the student's responsibility to officially withdraw if they do not plan to attend after registering. Students will not be automatically dropped for non-attendance. Repayment of certain types of financial aid administered through the University may be required as the result of dropping classes or withdrawing. For more information, contact the Office of Financial Aid and Scholarships (http://wweb.uta.edu/aao/fao/).

Title IX:

The University of Texas at Arlington is committed to upholding U.S. Federal Law "Title IX" such that no member of the UT Arlington community shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity. For more information, visit www.uta.edu/titleIX.

Academic Integrity:

Students enrolled in this course are expected to adhere to the UT Arlington Honor Code:
"*I pledge, on my honor, to uphold UT Arlington's tradition of academic integrity, a tradition that values hard work and honest effort in the pursuit of academic excellence.*
I promise that I will submit only work that I personally create or contribute to group collaborations, and I will appropriately reference any work from other sources. I will follow the highest standards of integrity and uphold the spirit of the Honor Code."

UT Arlington faculty members may employ the Honor Code as they see fit in their courses, including (but not limited to) having students acknowledge the honor code as part of an examination or requiring students to incorporate the honor code into any work submitted. Per UT System Regents' Rule 50101, §2.2, suspected violations of university's standards for academic integrity (including the Honor Code) will be referred to the Office of Student Conduct. Violators will be disciplined in accordance with University policy, which may result in the student's suspension or expulsion from the University.

Any student found guilty of academic dishonesty will receive a -100% for that work (project, exam, homework, etc.) as well as having the course grade lowered one full letter grade - in addition to any other penalties assessed (suspension, expulsion, probation). These and other applying UTA rules, will be strictly enforced. Any case of academic dishonesty will be treated in accordance with the UTA ***Handbook of Operating Procedures*** or the Judicial Affairs website at http://www2.uta.edu/discipline. If you do not understand this policy, it is your responsibility to obtain clarification or any additional information you may require
Students are allowed to discuss homework with classmates, but are **not** allowed to copy the solutions of others or share solutions with others. All work turned in for grading must be the student's own work.

Students will be required to sign an academic honesty letter to be kept with the instructor. Failing to provide with such a letter by census day will result in the respective students' withdrawal from the class.

Americans with Disabilities Act:

The University of Texas at Arlington is on record as being committed to both the spirit and letter of all federal equal opportunity legislation, including the Americans with Disabilities Act (ADA). All instructors at UT Arlington are required by law to provide "reasonable accommodations" to students with disabilities, so as not to discriminate on the basis of that disability. Any student requiring an accommodation for this course must provide the instructor with official documentation in the form of a letter certified by the staff in the Office for Students with Disabilities, University Hall 102. Only those students who have officially documented a need for an accommodation will have their request honored. Information regarding diagnostic criteria and policies for obtaining disability-based academic accommodations can be found at www.uta.edu/disability or by calling the Office for Students with Disabilities at (817) 272-3364.

Electronic Communication Policy

UT Arlington has adopted MavMail as its official means to communicate with students about important deadlines and events, as well as to transact university-related business regarding financial aid, tuition, grades, graduation, etc. All students are assigned a MavMail account and are responsible for checking the inbox regularly. There is no additional charge to students for using this account, which remains active even after graduation. Information about activating and using MavMail is available at http://www.uta.edu/oit/cs/email/mavmail.php.

Student Support Services:

UT Arlington provides a variety of resources and programs designed to help students develop academic skills, deal with personal situations, and better understand concepts and information related to their courses. Resources include tutoring, major-based learning centers, developmental education, advising and mentoring, personal counseling, and federally funded programs. For individualized referrals, students may visit the reception desk at University College (Ransom Hall), call the Maverick Resource Hotline at 817-272-6107, send a message to resources@uta.edu, or view the information at www.uta.edu/resources.

Student Feedback Survey:

At the end of each term, students enrolled in classes categorized as "lecture," "seminar," or "laboratory" shall be directed to complete an online Student Feedback Survey (SFS). Instructions on how to access the SFS for this course will be sent directly to each student through MavMail approximately 10 days before the end of the term. Each student's feedback enters the SFS database anonymously and is aggregated with that of other students enrolled in the course. UT Arlington's effort to solicit, gather, tabulate, and publish student feedback is required by state law; students are strongly urged to participate. For more information, visit http://www.uta.edu/sfs.

Final Review Week:

A period of five class days prior to the first day of final examinations in the long sessions shall be designated as Final Review Week. The purpose of this week is to allow students

sufficient time to prepare for final examinations. During this week, there shall be no scheduled activities such as required field trips or performances; and no instructor shall assign any themes, research problems or exercises of similar scope that have a completion date during or following this week unless specified in the class syllabus. During Final Review Week, an instructor shall not give any examinations constituting 10% or more of the final grade, except makeup tests and laboratory examinations. In addition, no instructor shall give any portion of the final examination during Final Review Week. During this week, classes are held as scheduled. In addition, instructors are not required to limit content to topics that have been previously covered; they may introduce new concepts as appropriate.

Emergency Exit Procedures:

Should we experience an emergency event that requires us to vacate the building, students should exit the room and move toward the nearest exits which are located both left (west) and right (east) after you exit Room 129ERB (a map with detailed red arrows showing the paths is located next to the classroom door). When exiting the building during an emergency, one should never take an elevator but should use the stairwells. Faculty members and instructional staff will assist students in selecting the safest route for evacuation and will make arrangements to assist handicapped individuals.

Wk.	Date	First half of class	Second half of class	Homework Announced (A) or Due (D)
1.	08/22/14	Introduction	Probabilities	
2.	08/29/14	More about Probabilities	Bayesian Examples	H1-A
3.	09/05/14	Multivariate μ and σ	Distributions	
4.	09/12/14	Distributions	Moment generating fun.	H1-D; H2-A
5.	09/19/14	Finish prob. Heavy tail	Intro to Information Theory	
6.	09/26/14	Info. Theory	Info. Theory	
7.	10/03/14	Random Numbers	Review	H2-D
8.	10/10/14	1. midterm	Random Numbers	H3-A
9.	10/17/14	Random Numbers	Statistics start	
10.	10/24/14	Statistics	Confidence intervals	H3-D
11.	10/31/14	Statistics & Tests	Statistics& Tests	
12.	11/07/14	Bayesian Networks	Bayesian Networks	H4-A
13.	11/14/14	Markov Models	Markov Models	
14.	11/21/14	2. midterm	Queuing Theory	H4-D, H5-A
+1	12/05/14			H5-D

As the instructor for this course, I reserve the right to adjust this schedule in any way I deem beneficial; this includes adding to or omitting topics listed above. – Gergely V. Záruba