**Computing Practice and Theory**

**Faculty: Judy Cushing, Aaron Skomra, Richard Weiss**

[**http://blogs.evergreen.edu/cpat**](http://blogs.evergreen.edu/cpat)

|  |  |  |  |
| --- | --- | --- | --- |
| Monday | Tuesday | Wednesday | Thursday |
| 10-12 Lecture,  Sem 2 E 1105 | 9:30-12 – Lab  Modeling, Stats  LIB 2617 & 2605 | 10-12 Lecture, Sem 2 E 1105 | 9:30-12 Lab –  Data Mining  LIB 2617 & 2605 |
| 1-3 Seminar – SEM II E 2107, 2109, 3107 | 1-3 Student Project Meetings with faculty Lib 2605 (AC Conf. Room), LIB 3509 & 3502 (Library proper)  Open Lab - Lib 2617 |  | 1-1:30: Program Mtg, Lib 1412  1:30-3 Student Project Meetings with faculty Lib 1412  Open Lab - Lib 2617 |
|  | 3:30-5 – Guest Lectures - LH 5 |  |  |

This project-oriented program for intermediate and advanced computer science students will weave together theory and practice in two cross-cutting computer science areas (pattern analysis and modeling) in the context of eScience.  The overriding question of the program is how pattern analysis and modeling advance the natural and physical sciences, particularly environmental science and climate change studies. We aim to consider in particular the promises and pitfalls of “Big Data”. The program aims to give students from *Computability*, *Computer Science Foundations*, and *Music, Math and Cybernetics* opportunities to continue work begun in those programs. Students who have taken *Computability* will be expected to complete more advanced work for upper division credit.

Particularly in seminar, students will share responsibility for presenting and discussing concepts from readings and lectures. The seminar will focus on how computation and statistics can have an impact on science and society through data mining, visualization, and machine learning. This program will include a guest lecture series that focuses on (how computers are used in) science.  In addition to seminar, the program has two disciplinary components (1) the theory and practice of statistics, and (2) data mining, machine learning and pattern recognition, and (3) a project.

For the project, which can be either a research project or a programming project, students will apply the computing and statistics to a problem of their choice.  Projects will begin with a proposal and bibliography, and should be either small enough in scope to be completed in one quarter or a self-contained part of a larger project.  CS subdisciplines where faculty will encourage projects include data mining, machine learning, statistics, database systems, data visualization (especially visual analytics), networking, security, algorithmic complexity, and formal languages, but they will entertain any reasonable proposal where learning objectives are clear and likely to be attained, resources are available to complete the project, the project scope is appropriate, and the project is linked to some area of computer science. To facilitate projects, faculty will organize small research groups that meet twice weekly (once with a faculty advisor).

Proposed credit distribution:

4:  Computing theory and practice: Advancing the Practice of Science (Seminar & Lecture).

4:  Modeling and Statistics

4:  Data Mining, Machine Learning and Pattern Recognition

4:  Student project - research and/or programming practicum

Upper Division Science Credit: Students seeking to earn upper division credit must submit to faculty by the end of the first week of the quarter responses to a questionnaire establishing the expertise and prior learning needed to complete work at an upper level.

Wednesday, April 24, No Class so students and faculty can observe and participate in Day of Absence.

**Books:**

Statistics: Nicholas Gotelli and Aaron Ellison, *A Primer of Ecological Statistics*, 2nd edition $54.95. Sinauer Associates, Inc.; 2nd edition (November 15, 2012), ISBN-13: 978-1605350646

Data Mining: Ian H. Witten , Eibe Frank , Mark A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd Edition (Morgan Kaufmann Series in Data Management Systems). 3rd edition (January 20, 2011), ISBN-13: 978-0123748560

Seminar Readings:

* Donella Meadows, *Thinking in Systems*. Chelsea Green Publishing (December 3, 2008).
* [David Salsburg](http://www.amazon.com/David-Salsburg/e/B001HPC4DI/ref=sr_ntt_srch_lnk_1?qid=1359323786&sr=1-1), [*The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*](http://www.amazon.com/Lady-Tasting-Tea-Statistics-Revolutionized/dp/0805071342/ref=sr_1_1?s=books&ie=UTF8&qid=1359323786&sr=1-1&keywords=lady+tasting+tea). Holt Paperbacks (May 1, 2002),
* Jaron Lanier, *You Are Not a Gadget: A Manifesto*. Vintage; Reprint edition (February 8, 2011),

\*Guest Lecture Series: In addition to the seminar books above, students will read articles related to guest lectures. Potential lecturers:

* Bill Howe, computer scientist, eScience Institute, University of Washington
* Veronika Megler, computer scientist, Portland State University
* Bob McKane, hydrology engineering, Environmental Protection Agency, Corvallis Oregon
* John Bolte, environmental engineering and land use planning, Oregon State University
* Denise Lach, science policy analyst, Oregon State University
* Nik Stephenson-Molnar, software engineer, Conservation Biology Institute, Corvallis, OR
* Dominique Bachelet, climate change impacts scientist, Oregon State University
* Jeremy Littel, climate change scientist, University of Washington
* Tom Dieterich, computer scientist, Oregon State University

Software:

* Statistics: JMP (available in the computing center)
* Modeling: Stella <http://www.iseesystems.com/softwares/Education/StellaSoftware.aspx> Proprietary software - isee systems. We will use this for less than two weeks of the quarter. It might be possible to download a trial version.
* Data Mining: Weka (Waikato Environment for Knowledge Analysis) <http://www.cs.waikato.ac.nz/ml/weka/>. available as a download – be sure to download the stable version (not the development version). Weka software toolkit downloadable with purchase of the text. A suite of open source machine learning software written in Java, developed at the University of Waikato. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; *SIGKDD Explorations*, Volume 11, Issue 1.