CSCI 568: Data Mining
Fall/Winter 2008, M/W/F 3-4PM, 134 Alderson Hall

Yong Joseph Bakos
205 Stratton Hall
(303) 653-3017
ybakos@mines.edu

Office hours are M/W/F from 4 – 5PM, or by appointment.

Course Web home: http://mines.humanoriented.com/classes/2008/fall/csci568/

Prerequisites: CSCI403 Database Management or approval of instructor, and a passion for programming.

Texts: Programming Collective Intelligence. Segaran. O'Reilly Media. 2007.
Introduction to Data Mining. Tan, Steinbach, Kumar. Addison-Wesley. 2006.

Course Objectives

The goal of this course is to understand and *to be able to programmatically apply* the basic concepts of data mining. Topics include:

- Data mining theory and algorithms
- Mathematical foundations of data mining tools
- Programming data mining algorithms
- Acquiring, parsing, filtering, mining, representing, refining and interacting with data
- Data visualization

At the end of this course you should have a complete data mining portfolio containing implementations in Python, and some cool data visualizations in Processing.

Grading

- Portfolio 50%
- Midterm 20%
- Final 20%
- Homework & quizzes 10%

Data Mining Portfolio

Most of your labor in this class will involve continuous work on your class portfolio. While each assignment is not large, it's **very important** to program regularly. In addition, the algorithms implemented programmatically will be constantly discussed in class, and it is critical to keep the pace up.

Exams

One midterm exam will be conducted the week of October 6 2008.

The final exam will be conducted during the week of December 8 2008.

A makeup examination can be arranged only when a student has an emergency (eg, medical emergency or urgent family matter). The student may be asked to provide the instructor with an appropriate document, such as a doctor's note.

Accommodation

If you need certain accommodation based on disability, talk to the instructor in person so that appropriate arrangements can be made.

Course Schedule

This schedule is not fixed in stone and is subject to change according to the actual progress of the course.

Week	Lecture	Reading
1	Introduction, Python	DM ch 1, CI ch 1
2	Python	Py tutorial, doco
3	Filtering, Similarity	CI ch 2, DM ch 2
4	Clustering	CI ch 3, DM ch 8
5	Matching and Ranking, Neural Nets	CI ch 4
6	Stochastic Optimization, Genetic Algorithms	CI ch 5
7	Classification	CI ch 6, DM ch 4
8	**Midterm**	
9	Decision Trees	CI ch 7, DM ch 4
10	Graphs, Neighbors	CI ch 8, DM ch 9
11	Advanced Classification, Support-Vector Machines	CI ch 9, DM ch 5
12	Features, Non-Negative Matrix Factorization	CI ch 10
13	Genetic Programming	CI ch 11
14	Visualization	DM ch 3, processing
15	Visualization	Processing doco
16	Algorithm Review	CI ch 12
17	Final Exam	

On Collaboration & Academic Integrity

Students are encouraged to discuss and collaborate as much as possible. However, it is obviously not acceptable to copy another student's solution. Your work must be your own. In addition, simply copying solutions found online is not acceptable. Be aware that homework assignments, project and midterm will not just focus on producing correct code, but explaining how things work.

Please see the Student Handbook for details on academic dishonesty. No exceptions will be made for students found simply giving away or taking another's solutions.

Academic Integrity Pledge

I pledge to uphold the high standards of academic ethics and integrity expressed by the Colorado School of Mines Student Honor Code by which I am bound. In particular, I will not misrepresent the work of others as my own, nor will I give or receive unauthorized assistance in the performance of academic coursework. I understand that my instructor will report any infraction of academic integrity to the Department Head and that any such matter will be investigated and prosecuted fully.

Name (print): _____

Signature: _____