COSC6335: Data Mining

Instructor: Carlos Ordonez

Course information
Schedule: Tu Th 11:30-13:00
Office hours: Tu Th 10:30-11:30
email: ordonez@cs (Start subject line with "COSC6335-")
TA: Zhibo Chen (zchen6@cs.uh.edu)

1 Course contents
This is a graduate level database course on data mining. The focus of the course is understanding how machine learning and statistical techniques work in a database system, especially for large data sets. The course has some overlap with machine learning, but it is less theoretical and more systems oriented. Notice the course is self-contained: it will cover the basics of relational DBMSs and SQL. The textbook is [2], complemented by [1, 3]. The course will also require reading some research papers.


2 Grading

- 70%: 8 assignments.
- 30%: Final exam.

All homeworks must be turned in to get B-. The course will use a relational DBMS and a modern data mining tool. Assignments will involve programming, using a data mining tool and writing some reports. Written reports are preferred in Latex/PDF, but any word processor is acceptable. Programs will be developed in Java and SQL (C++ or C# are not allowed). We will use scientific data sets for analysis. All assignments must be done in pairs (team of 2 students).

References
