

## *Data Mining*

**Instructor:** Dr. Joydeep Ghosh

**Course Description:** Many companies that gather huge amounts of electronic data have now begun applying data mining techniques to their data warehouses to discover and extract “hidden” patterns useful for making smart business decisions. Effective data mining requires an understanding of concepts from exploratory data analysis, pattern recognition, machine learning/ AI, heterogenous data bases, parallel processing and data visualization, in addition to knowing the application domain. I will focus on basic techniques for data mining, including methods useful for analyzing information from the world wide web. Demos using an industrial strength software (SAS) as well as a public domain JAVA package (WEKA) will be given and some applications/case studies will be discussed. The course involves a mid-term exam, a paper presentation and a term project. There will be no final exam.

**Syllabus:** This course requires students to have very basic knowledge of JAVA. An undergraduate level understanding of probability/statistics, data analysis, databases and linear algebra is assumed. This is a graduate course so the workload will be medium to heavy.

While studying techniques for database representation/modeling, clustering, classification, finding associations and sequence processing, emphasis will be placed on the issues of algorithm scalability, performance, interpretability and the ability to deal with garbage data. 10-15 minute student talks will be interwoven with the lectures, depending on class size. The last two classes will largely consist of student term-project presentations, followed by active discussion.