Foundation of Data Mining: Introduction

Hillol Kargupta
CSEE Department, UMBC
hillol@cs.umbc.edu
ITE 342, (410) 455-3972
www.cs.umbc.edu/~hillol

Acknowledgement: Tan, Steinbach, and Kumar provided some of the slides

Background

- Undergraduate level knowledge in
  - Linear Algebra
  - Probability Theory
  - Algorithm design and analysis
- Familiarity with a modern programming language.

Grading Policy

- Homework/Quizzes: 20%
- Exams: 45%
- Term project: 35%

Text Books

Primary Text:

- V. Kumar et al. (2005). Data Mining. Addison Wesley. (Not yet available in the market; we will use internally available copies available at the UMBC Bookstore).

References:

- J. Han and M. Kamber (2000) Data Mining: Concepts and Techniques, Morgan Kaufmann. Morgan Kaufmann Publishers; ISBN: 1558604898.
- Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations; by Ian H. Witten, Eibe Frank. Morgan Kaufmann Publishers; ISBN: 1-55860-552-5.
- T. Hastie, R. Tibshirani J. Friedman. (2001). The Elements of Statistical Learning: Data Mining, Inference, and Prediction.
- D. Hand, H. Mannila, P. Smyth (2000), Principles of Data Mining , MIT Press. S. M. Weiss and N. Indurkhya, Predictive Data Mining: A Practical Guide, Morgan Kaufmann Publishers, 1998. ISBN: 1-55860-403-0.
- Research papers

Software

- Weka: An open source data mining system
- Provides implementations of various data mining algorithms
- MATLAB, Mathematica

Evolution of Problems, Contexts, & Solutions

Cave Pictures

Advanced Word Processing

- Problems evolve
- Contexts change
- Calls for new perspectives and new approaches