Computer Science 477

Data Mining – Introduction and Expectations

Lecture 1

Syllabus

- How to contact me
- Course web site (Canvas)
- Course description
- Grading
- Academic dishonesty
- Topic Schedule

Teaching and Learning are Mutually Cooperative Ventures

- What you may expect of *me*:
  - To begin class on time (or *try*)
  - To be prepared for the lecture
  - To respond courteously to questions
  - To return homework and exams as soon as possible
  - To hold regular office hours
  - To make appointments with those that cannot attend office hours
  - To treat every student with equal fairness
  - To act as a professional at all times

Teaching and Learning are Mutually Cooperative Ventures

What I expect from *you*

- To be prepared for class by having read the assigned material
- To be courteous to your fellow classmates by not arriving late (or *try*), leaving early, or talking
- To be aware of all information in the course syllabus
- To be aware of all information on the Canvas site
- To ask questions when you do not understand, either in class, or at office hours
- To return exam or homework to me with written questions within 48 hours of the time the exam is returned to you if you have a question about the grading
- To not ask for special treatment

Organization of this class

- Tuesday, Wednesday lectures
- Attendance is *required*
  - Graded exercises in class
- Why daily exercises
  - Experience with students
  - Psychology / learning theory
- Spirit of the course:
  - Contribute to your professionalization
  - Contribute to your ability learn independently
  - Contribute to your ability work collaboratively
  - Contribute to your judgement about data, information

Requirements – Graded Work

- Exercises to be completed during lecture, or very soon thereafter   -   3-6 points each
- Homework   -   20-40 points each
- Project Progress Reports   -   10 points each
- Project Presentation   -   30 points
- Final Report   -   100 points
- Final Examination   -   100 points

Quarter Project

- (1) Preliminary exploration (using, perhaps, EDA, Descriptive Statistics) of possible datasets to explore

- (2) Project goal formulation, including choice of datamining technique(s) to apply,

- (3) Data preparation,

- (4) Conducting the project (which will include stumbles and backtracking),

- (5) Project presentation, and

- (6) Project write-up

What Data Mining Can Do

- Classification
  - Devise a predictive model
  - Configuration of attribute values → classification
  - I.e., Is a transaction with certain attribute values fraudulent?
- Clustering
  - Given (1) data set, (2) attribute values, (3) measures of similarity between items →
  - Natural groupings
- Text Mining
  - All of the above

What Data Mining Can Do

- **Association Rule Mining**
  - Given: a collection of collections
    - A collection of poems, which are a collection of words
    - Supermarket transactions, which are a collection of items purchased together
  - Determine what items appear in the same collection
    - Oreos and olives
    - "Moon" and "mudshark"
- **"Social Network" mining.**
  - Given a database of communities
  - Are their identifiable communities?
- **Sequence mining**
  - Characteristic ordering of events.

Right Away

- Form Project Teams

- Identify promising datasets, research questions

- First progress report due Monday, April 4