

36-350 DATA MINING

Fall 2007

Instructor: William F. Eddy

Office: Department of Statistics, BH132F

Phone: 268-2725

Email: bill@stat.cmu.edu

Grader: Daniel Manrique

Office:

Phone:

Email: danielm@stat.cmu.edu

Class Meetings: Monday and Wednesday PH 226B 10:30-11:20am

Lab: Friday BH140E 10:30-11:20

Office Hours: I will hold my “official” office hours immediately following class on Monday and Wednesday. I am generally available in my Baker Hall office all day every day except Friday or when otherwise occupied.

Course objectives:

1. Learn some basic tools of exploratory data analysis, including graphical displays and methods of reexpression;
2. Learn the language of data mining and learn about the structure of various kinds of data;
3. Learn about measures of information and methods of clustering;
4. Learn a little about graphical methods and visualization;
5. Learn the basics of linear models and logistic regression.

Course work: Your grade in this course will be determined by homework assignments, labs, and a final project. There will be weekly homework assignments handed out on Wednesday and due the following Wednesday.

The project will involve the analysis of a real data set and will require a written report. The written report will include two sections, one a reasonably technical one (of the sort I would expect a consultant to write) and the second a non-technical interpretation (of the sort that might be presented to a non-technical manager). OUTPUT FROM A COMPUTER PROGRAM IS NOT ACCEPTABLE (even for the technical part). The project will be assigned later in the semester.