Statistical Data Mining

Fall 2011

Instructor Information: Dr. Alan M. Polansky
Office: Du Sable 359D
Hours:
E-Mail: polansky@math.niu.edu


Grading: Your grade in this course will be based on assigned homework (10% total), three small projects (20% each), and a final project (30%)

Homework: Homework will be assigned daily, with due dates being given in class. **It is the responsibility of the student to keep track of which assignments are due each day.** No late assignments will be accepted for any reason. **Assignments are to be neat and presented logically on loose-leaf paper.** If more than one sheet of paper is used, the assignment must be stapled together. Please do not ask faculty for use of a stapler. **Spiral paper is unacceptable.** You are to show all your work. **If your assignments are messy, disorganized or do not follow the previously stated instructions, you will not receive credit for that assignment.** Please note that mistakes in your grades on homework, and exams may occur. You have one week from the time the paper was handed back to notify me of the mistake. **If you wait longer than the one week, no change will be made to the grade.**

Incomplete Grades: Please note that a grade of incomplete (I) will only be considered for students who are passing the course, but cannot complete the course due to health or family reasons. **A grade of incomplete will not be assigned to anyone who is not passing the course at the time of the request.**

Extra Credit: **There is no extra credit given in this course. Don’t even ask!**

Attendance: There is no mandatory attendance requirement for this class. **Please note, however, that if you frequently miss class the instructor reserves the right to deny office hour privileges to you.** Further, it is your responsibility to be sure that your work is turned in and that you get the notes and announcements, if any, from a student in the class. My notes will not be available to students.

Syllabus Changes: The instructor reserves the right to amend the syllabus at any time. Changes will be announced in lecture.

FERPA: Due to privacy concerns, I will not discuss grades using email or the telephone. You must meet with me in person to discuss your grades.

Academic Misconduct: All university regulations regarding academic misconduct will be followed in this class. Evidence of academic misconduct will be forwarded to the Northern Illinois University judicial system and may result in a failing grade for the course or expulsion from the university.

CAAR: This university abides by Section 504 of the Rehabilitation Act of 1973, which stipulates that no student shall be denied the benefits of an education solely by reason of handicap. If you have a disability that may have some impact on your work in this class and for which you may require accommodations, please see me and the staff at the Center for Access-Ability Resources (CAAR) located at the Health Services Building so that accommodations may be arranged.
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Topics</th>
</tr>
</thead>
</table>
| 1       | Introduction to Data Mining  
          Introduction to Machine Learning |
| 4       | Density Estimation  
          Statistical Properties of Density Estimators  
          The Histogram  
          The Kernel Density Estimator |
| 8       | Linear Discriminant Analysis  
          Binary Classification  
          Multi-class Problems |
| 9       | Classification Trees  
          Regression Trees |
| 11      | Linear and Nonlinear Support Vector Machines  
          Multi-class Problems |
| 12      | Cluster Analysis  
          Hierarchical Clustering  
          Partitioning Methods  
          Self Organizing Maps  
          Clustering Variables  
          Block Clustering  
          Clustering Microarray Data  
          Mixture Models |
| 14      | Bagging and Boosting |