DATA MINING
CSCI-B490
Fall 2015

BASIC INFORMATION

Class meets:
Time: TR 5:45pm – 7:00pm
Place: Informatics West 107

Instructor:
Predrag Radivojac
Office: Lindley Hall 301F
Email: predrag@indiana.edu
Web: www.cs.indiana.edu/~predrag

Office Hours:
Time: Tuesdays 11:00am-12:00pm
Time: Thursdays 2:00pm-3:00pm
Place: Lindley Hall 301F

Course Web Site:
http://www.cs.indiana.edu/~predrag/classes/b490.htm

BASIC INFORMATION

Associate Instructor:

Hasan Kurban
Email: hakurban@indiana.edu
Office: Lindley Hall 310

Office Hours:
Time: Mondays and Wednesdays, 11:00am-12:30pm
Place: Lindley Hall 310
TEXTBOOK

Introduction to Data Mining - by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar

Chapter 1: Introduction
Chapter 2: Data
Chapter 3: Exploring data
Chapter 4: Classification
Chapter 6: Association analysis
Chapter 8: Cluster analysis
Chapter 10: Anomaly detection

Supplementary material will be provided in class!

ALSO GOOD READINGS...

• Data Mining: Concepts and Techniques - by Jiawei Han and Micheline Kamber

• Data Mining: Practical Machine Learning Tools and Techniques - by Ian Witten and Eibe Frank

• Principles of Data Mining - by David Hand, Heikki Mannila, and Padhraic Smyth

Lecture Slides

• Introduction to Data Mining - by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar

• Data Mining: Concepts and Techniques - by Jiawei Han and Micheline Kamber

• Summary: Our own slides + some mix from the slides for the books above
OVERVIEW OF THE COURSE

See online syllabus…

- introduction to data mining
- introduction to Matlab and Matlab programming
- data representation and data preprocessing
- data visualization
- prediction methods (classification and regression)
- mining association rules
- clustering techniques
- privacy-preserving data mining
- case studies on various types of data
- and more… (how much? we’ll see!)
GOALS

• This course is designed to introduce basic concepts of Data Mining and provide hands-on experience to data analysis, clustering, and prediction using Matlab.
• The students will be expected to develop a basic understanding of Data Mining and develop skills to solve practical problems.
• Data Mining is a practical discipline that aims to identify interesting new relationships and patterns hidden in numerous databases and real life.

HIDDEN GOALS

• To appreciate fundamental mathematical concepts
• To appreciate abstraction and to not be afraid of it
• To be able to understand how to transfer solutions from one set of problems to another set of problems (through abstraction)
• To learn to be a problem solver
• To recruit undergraduate researchers
• NSF Graduate Research Fellowship Program

WHAT DO I ASSUME?

• You have basic programming skills
• You have basic mathematical skills (e.g. calculus)
• You are patient

WHAT WOULD I LIKE TO SEE?

• You are motivated to learn
• You are motivated to succeed in this class
**GRADING**

- Midterm exam: 20%
- Final project (groups are OK): 30%
- Homework assignments (5-6): 30%
- Class participation and quizzes (4): 20%

- all assignments are individual
- top performers in the class will earn A
- distributions of scores will be generated after assignments; regularly
- you will know where you stand in the class, if you don’t - ask me
- do not expect late I’s or W’s

---

**LATE ASSIGNMENT POLICY**

- The homework assignments are due on the specified due date through Oncourse
- Late assignments will be accepted (unless there are legitimate circumstances) using the following rules

<table>
<thead>
<tr>
<th>points x 1.0</th>
<th>(on time)</th>
<th>recommended!</th>
</tr>
</thead>
<tbody>
<tr>
<td>points x 0.9</td>
<td>(1 day late)</td>
<td>} not recommended!</td>
</tr>
<tr>
<td>points x 0.7</td>
<td>(2 days late)</td>
<td>} not recommended!</td>
</tr>
<tr>
<td>points x 0.5</td>
<td>(3 days late)</td>
<td></td>
</tr>
<tr>
<td>points x 0.3</td>
<td>(4 days late)</td>
<td></td>
</tr>
<tr>
<td>points x 0.1</td>
<td>(5 days late)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>(after 5 days)</td>
<td></td>
</tr>
</tbody>
</table>

---

**ANNOUNCEMENTS**

- Midterm exam – wk 7 (in class); October 8
- Quizzes (4!) – wk 4, wk 9, wk 11 (all in class)
- Final presentations – Last week of classes
  December 8 and 10th
- Labor day – September 7 (no class)
- Thanksgiving break – November 21-29 (no classes)
ACADEMIC HONESTY

- Code of Student Rights, Responsibilities, and Conduct !!!
  - [http://www.indiana.edu/~code/](http://www.indiana.edu/~code/)
  - Many interesting things there, including that... Students are responsible to "facilitate the learning environment and the process of learning, including attending class regularly, completing class assignments, and coming to class prepared."

- Academic honesty taken seriously!
  - I am obliged to report every cheating incident to the university
  - Do the right thing

MISCELLANEA

- Do not record instructor(s) without explicit permission
- Turn off cell phones and other similar devices during class.
- Use laptops if you have to (unless it bothers someone)
- "will u be in ur office after class?"; "I need a letter of recommendation."

- BE NICE TO PEOPLE