

Syllabus

Advanced Data Mining

Course Name	Course type (credit/hours)	전선(3/3)			Course code	
	Target students Division/major/grade	산업공학과/6학년			Opening semester	2017년 2학기
	Class time and classroom	목10(팔323) 목11(팔323) 목12(팔323)(팔323)				
Reference to this course	Related basic courses					
	Recommended concurrent courses					
	Related advanced courses	Machine Learning				
Instructor	Name (title/division)	신현정 (교수/산업공학과, 데이터사이언스학과)				
	Office Room Number	Padal 818	Office phone Number	2417	e-mail	shin@ajou.ac.kr
	Office hours	Wed 14:00~16:00		Homepage address	www.alphaminers.net	
Teaching Assistant	Name (title/division)					
	Office Room Number	Paladal 823	Office phone Number	1866	e-mail	ldg1226@ajou.ac.kr

1. Introduction

Data capture has become inexpensive and ubiquitous as a by-product of innovations such as the internet, e-commerce, electronic banking, point-of-sale devices, bar-code readers, intelligent machines, and the amount has been increasing at an incredible rate due to technological advances. "Data mining" refers to a collection of techniques for extracting "interesting" relationships and knowledge hidden in a mountain of data in order to assist managers or analysts to make intelligent use of them. A number of successful applications have been reported in areas such as credit rating, fraud detection, database marketing, customer relationship management, and stock market investments. In this course, we will examine a variety of data mining techniques evolved from the disciplines of statistics and artificial intelligence (or machine learning), and practice them in recognizing patterns and making predictions from an applications perspective. Application (or case) surveys and hands-on experimentations with easy-to-use software will be provided.

2. Course Objectives

The course is aimed at providing students with the most important fundamentals and techniques of Data Mining:

- Predictive modeling: Classification, Regression (decision trees, neural networks, statistical models, etc)
- Association rules and Link analysis
- Clustering
- Anomaly Detection
- Interpretations
- Some Advanced Machine Learning techniques
- etc.

With presentation of the corresponding theories, the course still maintains its practical approach with statistical S/W packages (e.g., SAS E-miner/MATLAB/MINITAB, etc)

3. Class types and activities

4. Teaching Method

[Lectures]

The class introduces basic theory of data mining techniques. For student's better comprehension, there will be assigned readings to be done before class, and be shortly checked every beginning of class by questioning or randomly selected individuals. The students will practice what they learn by solving the assigned homework, very often with S/W packages.

[Practice Course]

SAS Enterprise Miner will be available for projects that require handling large amounts of data. Instructions on using the software will be provided in recitations during the course.

[Case/Survey write-ups]

Occasionally, readings for more advanced topics and cases will be suggested to augment the lecture notes.

[Term Project]

The term project is a group project intended to tie together concepts and techniques discussed in class. You are encouraged to work together in pairs. Your group will write a project proposal, present the midterm/final presentation, carry out the proposed project, and submit a final project report of no more than 15 pages in length. The goal of the project is to produce a report that could be published in a regional conference related to your major and data mining. More details about the project and the major milestones for the project will be announced during the course.

5. Knowledge and ability required for taking this course

6. Method of Evaluation

Evaluation Item	The Number of Times	Evaluation Proportion	Remarks
Attendance			
midterm exam			

6. Method of Evaluation

Evaluation Item	The Number of Times	Evaluation Proportion	Remarks
final exam			
quiz			
presentation			
discussion			
homework			
etc			

[Grading]

The course grade will be based on the following four components. The numbers in the parentheses indicate the weights associated with the components: Four sets of Quizzes (40%), HW Assignment(30%), Term–Project(30%), and Attendance(Pass/Fail). (Note: According to instructor's decision, actual scores can be scaled to the (absolute) evaluation measurement.)

[Testing]

All tests will be given during the regular class sessions. No makeup test will be given under any circumstance. An excuse given to the instructor is only acceptable with proof–documents(e.g., a note from a physician in the case of illness)

[Homework Assignment and policy]

Each assignment will have a specific time for which it will be due. An assignment turned in late within three days of the due time will be docked 70%. No late submissions will be accepted after three days when it is due.

[Attendance]

Students are expected to attend all scheduled classes. It is the student's responsibility to find out what was discussed in a missed class. Although, attendance records will not be used to compute grades (except possibly in borderline cases), however, missing classes can be expected to significantly reduce the chances of success. "Failure of the course (F)" : if more than four times (unexpected) absence. (also, note that 2 times late = 1 unexcused absence)

7. Textbooks

Main/Sub	Title	Writer	Publisher	Publication year
주교재	Introduction to Data Mining	Tan, Steinbach, & Kumar	Pearson–Addison Wesley	2006
부교재	Data Mining: Practical Machine Learning and Techniques (2nd eds)	Witten & Frank	Elsevier	2005
부교재	Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management (2nd eds)	Berry & Linoff	Wiley	2004
부교재	Getting Started With Enterprise Miner Software 5.2	SAS Institute	SAS Institute	2006
참고자료	Pattern Recognition & Machine Learning	Bishop	Springer	2007

7. Textbooks

Main/Sub	Title	Writer	Publisher	Publication year
참고자료	Pattern Classification (2nd eds)	Duda, Hart, & Stork	Wiley-Interscience	2000

8. Lecture Schedule

Week	Lecture contents	Lesson type	Remark
1	[Course Introduction]	Theory	
2	[Exploring Data]	Theory/Practice/HW	
3	[Classification and Regression Techniques] - Basic Concepts - Decision Trees	Theory	
4	[Quiz1]	Quiz1	
5	[Classification and Regression Techniques] - Rule-based Classifier - Nearest-Neighbor Classifiers - Neural Network	Theory/Practice/HW	
6	[Classification and Regression Techniques] - Ensemble Methods - Support Vector Machine	Theory/Practice/HW	
7	[Quiz2]	Quiz2	
8	[Mid-term Presentation]	Students' Presentation	
9	[Association Analysis] - Basic Concepts and Algorithms	Theory	
10	[Association Analysis] - Advanced Concepts	Theory/Practice/HW	
11	[Quiz3]	Quiz3	
12	[Cluster Analysis] - Basic Concepts and Algorithms	Theory	
13	[Cluster Analysis] - Additional Issues and Algorithms	Theory/Practice/HW	
14	[Anomaly Detection]	Theory	
15	[Quiz4]	Quiz4	
16	[Term Project Presentation]	Students' Presentation	

9. Others

- * General

- (1) All tests will be given during the regular class sessions.
- (2) No makeup test will be given under any circumstance.
- (3) An excuse given to the instructor is only acceptable with proof—documents (e.g., a note from a physician in the case of illness)

- * There are about six sets of homework assignments.

- (1) Due dates for the submissions will be announced in class.
- (2) Late submissions within three days from when it is due will be based on 70% from the full credit when it is due.
- (3) No late submissions will be accepted after three days when it is due.

- * Students are expected to attend all scheduled classes.

- (1) It is the student's responsibility to find out what was discussed in a missed class.
- (2) Although, attendance records will not be significant to compute grades (except possibly in borderline cases), however, missing classes can be expected to "significantly" reduce the chances of success.

The course assumes only a modest statistics or mathematics background, and no database knowledge is needed. However, it is "highly" recommended for students to acquire basic concepts and techniques on data analysis from "Data Analysis and Practice" (undergraduate course, open every 2nd semester, dept of industrial & information systems engineering) as a pre-requisite. Students not meeting these requirements

- * Cell phone use (or PDA) is strictly prohibited during lecture.
- * Cheating or report plagiarism will greatly affect your credit when detected.