

CSC 7442 – Data Mining and Knowledge Discovery from Datasets

Fall 2009

Louisiana State University Department of Computer Science

Instructor: Professor Evangelos Triantaphyllou
Office hours: Tuesday and Thursday from 1:30 to 3:00 PM or by appointment

Office: 164-D Coates Hall, Phone: 578-1348,
E-mail: trianta@lsu.edu
Web page: <http://www.csc.lsu.edu/trianta>

Classroom: 129 Tureaud Hall
Meeting Times: Monday and Wednesday: 3:40 PM - 5:00 PM.

Textbook: **Introduction to Data Mining**, by Tan, Steinback and Kumar, Addison Wesley, latest edition, and other materials to be given by the instructor.

Prerequisites:

Knowledge of a general purpose programming language, elements of graphs, and Boolean algebra.

GOAL:

The main goal of this course is to introduce the students to the fundamental issues of the fast emerging field of data mining and knowledge discovery from databases. Emphasis is given on the principles that govern this new area of computer science.

MAIN TOPICS:

1. Introduction

- Introduction to Data Mining and Knowledge Discovery in Databases
- The concept of data mining and knowledge discovery in databases
- The data mining tasks
- Relationships of DM/KDD with other research fields: statistics, machine learning

2. The Data

- Data pre-processing: data cleaning, scaling and transformations
- Dimensionality reduction, feature selection

3. Cluster Analysis

- K-means and hierarchical clustering

4. Other Classification Techniques

- Decision trees
- Support vector machines
- Nearest neighbor classifier, Bayesian classifier and ANN
- Model overfitting

5. Boolean functions based Classification Techniques

6. Guided Learning

7. Association Analysis

- The basic association rule learning algorithms

8. Model Evaluation

9. Some Special Topics on Data Mining and Knowledge Discovery in Databases

GRADING:	Homework assignments	: 35%
	1 Midterm Exam	: 15%
	1 Small project	: 20%
	1 Final Exam	: 30% ; this is comprehensive and closed books/notes.
	Total points	: <u>100%</u>

COMMENTS:

[1]:

If a student earns 100-90 points, an A grade will be guaranteed.

If a student earns 89-80 points, a B grade will be guaranteed.

If a student earns 79-70 points, a C grade will be guaranteed.

If a student earns 69-60 points, a D grade will be guaranteed.

An F grade will be earned for fewer points. (Note that a D grade is not a passing grade for graduate students.)

However, students may earn an A or B or C even below the previous thresholds if the final grades of the entire class depict a clustering pattern as discussed in the first day of classes.

[2]:

Class attendance and active participation are **strongly** encouraged (but not considered in the grade).

[3]:

No cheating will be tolerated.

[4]:

No wireless gadgets (including cell phones) are allowed any time (i.e., lecture and/or test) during this course.

Exceptions might be possible in special cases and only after prior permission.

Some instructions in preparing HWs and Exams

- 1) *Make sure that you understand a given problem before you attempt to answer it.*
- 2) *Try your best in solving **all** problems in a given HW or exam by **YOURSELF**.*
- 3) *Pay extra attention in the **organization** of your answers. Solutions should not be mixed up with scratch material in deriving them.*
- 4) *Present and staple your solutions **sequentially**.*
- 5) *Make sure that your English is correct, clear and easy to read. Although this is not a composition course, it is very important that you present your solutions in a highly professional and scientific manner.*
- 6) *If you get confused, do not worry. Just relax, organize your thoughts, and try to see the problem from a simple, but still accurate, point of view.*
- 7) *You should have no doubt that **you** (assuming you have studied adequately) are capable of answering all the problems in the HWs and exams!*
- 7) *New policy for “**wireless communication devices and hand held computers**”: All such devices must be turned off in all class meetings, including examinations, and kept out of your reach at all time. During exams only the most elementary calculators (with the basic arithmetic functions) are allowed. No advanced calculators will be allowed. It is **your responsibility** to come with a simple calculator for the exams.*

A hint...

Next is the first question that is very likely to be given to you in the mid term and final exams and the correct solution. It is usually assigned 5 points.

PROBLEM 1: (5 points)

What is the single most important step in solving any science / engineering problem?

Answer: To define the problem **correctly** and **enjoy** the solution process.

(You must underline the right words to earn all 5 points.)