<div align="center">

Course Syllabus
CS535/435 Introduction to Data Mining
Fall 2016

</div>

Instructor

Instructor: Lei Yu
Office: Q05, Engineering Building
Lecture hours: 2:50pm – 4:15pm (TR)
Lecture classroom: SW 112
Office hours: 1:30pm – 2:30pm TR, or by appointment
Email: lyu@cs.binghamton.edu
Webpage: http://www.cs.binghamton.edu/~lyu

Course Description

This data mining course introduces the concepts, algorithms, techniques, and applications of data mining. Topics include background of data mining, data preprocessing, classification, clustering, association-rules mining.

Prerequisites

CS 333 Algorithms
MATH 327 Probability with Statistical Methods

Course Objectives

Upon completion of this course, students will:
- have knowledge of basic concepts, algorithms, techniques, and applications of data mining;
- have proficiency in the use and application of probability and statistics theories and algorithm analysis techniques in understanding data mining algorithms;
- be familiar with data mining tools.

Textbook

- *Introduction to Data Mining*, Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Addison-Wesley, April 2005.

This is a highly recommended textbook. The library has two copies on course reserve. Three core chapters of this book are publicly available online.
http://www-users.cs.umn.edu/~kumar/dmbook/index.php

Main Topics

- Background of knowledge discovery and data mining
- Data preprocessing (e.g., data cleaning, transformation, dimensionality reduction, instance selection)
- Classification (e.g., decision trees, Bayesian classifiers, instance-based classifiers, rule-based classifiers, artificial neural networks, support vector machines, ensemble methods)
- Clustering (e.g., K-means, hierarchical clustering, density-based clustering)
- Mining association rules (e.g., Apriori, FP-growth)

Lecture Notes

Lecture Notes for class will be posted on blackboard (generally before lectures). *Lecture notes do not substitute for class attendance*, since (i) they will not be complete and (ii) significant parts of lectures, including discussions and in-class exercises, may not come from the class notes.

Grading

1. For **CS535** students, your grade will be based on

   - Four homework assignments   20% (5% each)
   - Two exams   40% (20% each)
   - 6-8 short quizzes   10% (2% each, counting the top 5 highest grades)
   - Presentation   15%
   - Project and Report   15%

   For **CS435** students, your grade will be based on

   - Four homework assignments   24% (6% each)
   - Two exams   50% (25% each)
   - 6-8 short quizzes   10% (2% each, counting the top 5 highest grades)
   - Presentation   16%
   - Project and Report (optional)   10 bonus points (at maximum) added to the base grade

2. Assignments There will be 4 assignments in the form of written exercises on key concepts and algorithms. They are developed in order to help you understand the lecture materials and prepare for the two exams.

3. Exams. Exams will be in class, closed notes, and closed book, unless otherwise specified (unlikely). Exam I will take place during (or very close to) Week 7, and Exam II during or close to Week 12. Exam II will cover the materials that were not covered in Exam I.

   You must take each Exam during the one time that we give it; we will not give makeup Exams.

4. Quizzes. The quizzes will take place during classes, and we will announce some of the quizzes at least one class period in advance in class. Some other quizzes will be pop-up quizzes (given right after the introduction of a topic). We will only consider the top five scores in deciding your overall quiz score. A discussion will be held after each quiz.

   You must take each quiz during the one time that we give it; we will not give makeup quizzes.

5. Group Presentation

   Each student will be required to form a presentation group (of 2 and only 2 students) to give one presentation on a selected topic (a list of topics provided by the instructor). The presenters will also be responsible for conducting group discussions and answering questions. Presentations will be held during the last 3-4 weeks of classes.

6. Group Project (required for graduate students)

   Each CS535 student will be required to form a project group (2-3 students) to conduct a term project on a topic proposed by the group and approved by the instructor. A project group can take either of the two general types of project: (1) data mining application - applying data mining algorithms and

techniques learned in this class to a real-world problem or a research problem in another field; or (2) data mining case study – investigating and analyzing existing mining applications in a given domain (e.g., business, healthcare, industry, etc.) and identifying potential opportunities for new data mining applications in the same domain.

Each project group is required to submit and present in class a project progress report (in presentation slides format) during the middle of the semester, and submit a final project report (in research paper format) at the end of the semester.

Academic Honesty Expectations

Please review the academic honesty document and make sure that you understand it! The link is at: http://www.binghamton.edu/watson/about/honesty-policy.pdf.

Discussion of general concepts and questions concerning the homework assignments among students is encouraged. However, each of you is expected to work on the homework solutions on your own. Sharing of any part of solutions is prohibited. If you are unclear about the policy, please consult with the instructor before you act.

Late Policy:

Each assignment is due at the beginning of class on the due date. Any assignment received within the next 24 hours will be penalized by 20% of the full credit; any assignment received within the time between 24 hours and 48 hours pass the deadline is penalized by 50% of the full credit; No assignment will be accepted after 48 hours pass the deadline. Rare exceptions of this policy may be made at the discretion of the instructor under demonstrably circumstances.