

Syllabus of CSC 240/440 Spring 2017: Data Mining

January 19, 2017

1 Course Objective

This courses will introduce fundamental concepts and techniques of data mining, including data attributes, data visualization, data preprocessing, classification methods, cluster analysis, and mining frequent patterns, association and correlation. Advanced topics could be included if the time allows, such as sparse learning, semi-supervised learning, and outlier detection.

Graduate students (CSC440) may receive additional questions or requirements in homework or projects.

2 Basic Information

- Instructor: Professor Ji Liu
- Time: TR 4:50 - 6:05pm
- Classroom: GRGEN 108
- Office Hour: W 4-6pm or by appointment
- Email: ji.liu.uwisc@gmail.com
- Course website: <http://www.cs.rochester.edu/u/jliu/CSC-240-440-2017spring.html>
- TA office hour:
 - Haichuan Yang (Mon. 2-4pm, CSB 725, hyang1990.cs@gmail.com),
 - Chuyang Ke (Tue. 3-4pm, CSB 725, cke@u.rochester.edu)
 - Chester Holtz (Thu. 10-11am, CSB 603, choltz2@u.rochester.edu)

3 Textbooks

- (Required) Introduction to Data Mining, Pang-Ning Tan, Michael Steinbach, Vipin Kumar
- (Recommended) Data Mining: Concepts and Techniques, 3rd edition, Jiawei Han, Micheline, Kamber, and Jian Pei

4 Topics

- Data preprocessing
- Clustering
- Data visualization / dimension reduction
- Classification
- Mining frequent patterns
- Association rule
- Semi-supervised learning
- Outlier detection

5 Point Distribution and Grading Scheme (Total points: 100)

Point Distribution:

- Homework: 8×5 homework
- Mid-term: 20
- Project: 30
- Attendance / class participation: 10

Grading Scheme:

- A: ≥ 90 or above top 15%
- A-: ≥ 85 or above top 25%
- B+: ≥ 80 or above top 40%
- B: ≥ 75 or above top 50%
- C: ≥ 70 or above top 70%
- D: ≥ 60
- E: < 60

*Some bonus points could be earned in this course, for example, homework may include bonus points.

6 Late Policies

All assignments are due at the beginning of class on the due date. A total of two (2) free late days may be used throughout the semester without penalty. You can choose to use zero, one, or two free late days toward any one late homework, depending on availability. You must specify on the first page of your late homework how many free late days to use.

The lecturer and TAs cannot accept any late submission if the submission does not satisfy the late policy.

7 Pre-requisite

- CSC 171
- CSC 172
- MTH 161
- One of the following is required: CSC 242 or CSC 262 or MTH 165
- Prerequisites will be strictly enforced.

8 Project

The course will select a data competition (e.g., from Kaggle) as the course project¹. All teams will work on the same data competition project. Each team may have up to three members. The team diversity is strongly encouraged. One bonus point can be earned if three members are with different majors, or the team includes both undergraduate and graduate member. Your project also include a presentation, whose weight is about 20 percent of this project. Your grade of the project will be measured from the following aspects

- Performance of your algorithm, e.g., the rank of your result in Kaggle or in this class
- The novelty of the algorithms you developed for the data competition

More grading details will be available when the project is released.

9 Academic Honesty Policy (steal from ECE 450)

The University of Rochester academic honesty policy applies to all assignments and exams for this class. The full-text of the academic honesty policy can be found at: <http://www.rochester.edu/College/CCAS/AdviserHandbook/AcadHonesty.html> In addition to the general guidelines mentioned in the above policy, for this course I require that: In examinations, you must work individually with no communication with others and use only materials/tools that have been explicitly allowed. For homework, you may discuss problems with your colleagues but final solutions need to be worked out, written and submitted individually. Any external material used should be clearly

¹<https://www.kaggle.com/competitions>

cited. In your own writings (example term papers, homework solutions, proposals etc), no more than one or two sentences may be used verbatim from any source.

READ THESE INSTRUCTIONS CAREFULLY! If any aspect of the academic honesty policy and guidelines for this course are unclear, please ask me for clarifications. Lack of awareness or understanding of this policy will not be an acceptable excuse or defense against disciplinary action.