

CSE4334/5334 Data Mining

1 Logistics

Chengkai Li

Department of Computer Science and Engineering
University of Texas at Arlington
Fall 2018



Self Introduction



Chengkai Li

- Associate Professor, CSE
- Office: ERB 628
- (817) 272-0162
- cli@uta.edu
- <http://ranger.uta.edu/~cli>

Research Interests

- Big data and data science (database, data mining, natural language processing)

Basics



Lectures

- Tue/Thu 12:30-1:50pm, ERB 130

TA

- Fatma Arslan fatma [DOT] arslan [AT] uta [DOT] edu

Instructor Office hours

- Tue/Thu 2:00-3:00pm, ERB 628

TA Office hours

- Wed 3:00-5:00pm, ERB 508

Preparation/Expectation



- ❖ Be hands-on and have good programming experience
 - Multiple significant programming assignments
 - You are expected to use Python
- ❖ Be comfortable with topics in your math, statistics, probability courses
- ❖ Expect heavy workload, challenging assignments, exams
 - Be hard-working; expect to spend many hours; likely your heaviest course.
 - Exam is demanding and comprehensive; almost no student can finish all exam questions.
- ❖ Plagiarism is absolutely not tolerated. No excuse or second chance.

Academic Integrity



Violations

- Cheating on test/assignment; Plagiarism; Collusion

Can I refer to external materials?

- Yes, but in your homework, source code, and documentation you must explicitly acknowledge the source of information.
- If you copy sentences (completely or partially) from other places, you must enclose them with quotation marks, in addition to provide references to the information source.
- Even if you rephrase, you still need to acknowledge the source.
- If you copy source codes (completely or partially) from other places, you must provide references to the information source.



What types of discussions are allowed?

- You can discuss topics related to assignments with your fellow students.
- But you cannot discuss your solutions.
- You must not provide your work (email, hard copy, or in any form) to anyone for any purpose. Following actions are not acceptable:
 - “I emailed it to my roommate/friend so that I can submit from their computer, since I couldn’t get online from mine.”
 - “I sent it to my roommate/friend so that I can compile and test my program on their computer, since mine was down.”

Academic Integrity



Tutorial: <http://library.uta.edu/plagiarism/index.php>

More information at <http://www.uta.edu/conduct/academic-integrity/index.php>

The chance of being caught is large; we use tools to diligently check and compare the documents and source codes that you submit to us.

The consequence is certain:

- I will submit the form of “faculty referral of honor code violation” to the university. No exception!
- Academic penalty in the context of this course: 0 on assignment/exam, reduced grade, failing grade of the course
- Penalty by the university: probation, suspension, expulsion, ...



- (Required) Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. Introduction to Data Mining. (Sample chapters at <http://www-users.cs.umn.edu/~kumar/dmbook/index.php>)
- (Required for relevant chapters) Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze. Introduction to Information Retrieval. (Free book at <http://nlp.stanford.edu/IR-book/>)
- (Reference) Jure Leskovec, Anand Rajaraman and Jeff Ullman. Mining of Massive Datasets. (Free book at <http://www.mmds.org/#ver21>)
- (Reference) Jiawei Han, Micheline Kamber and Jian Pei. Data Mining: Concepts and Techniques.
- (Reference) Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani. An Introduction to Statistical Learning with Applications in R. (Free book at <http://www-bcf.usc.edu/~gareth/ISL/index.html>)
- (Reference) I. H. Witten and E. Frank. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.

The Slides



The slides highlight the gist of most important concepts and techniques.

- It is not meant to be complete. Details may not be included.
- It may be simplified for ease of explanation.

Only studying the slides is not enough.

Many lecture notes are adopted from

- Vipin Kumar (Minnesota)
- Jiawei Han (Illinois)

Tentative Grading Scheme



- ❖ Homework (HW) 0% (No homework)
- ❖ Programming Assignments (P) 40% (Must be done independently)
- ❖ Pop quizzes (Q) 30%
- ❖ Final 30%
- ❖ You are required to attend classes and actively participate in discussions.
- ❖ All assignments must be electronically prepared. We won't accept images of handwritten answers and hand-drawn pictures.
- ❖ **Final Exam:** December 11th, Tuesday, 11am-1:30pm. ERB 130
- ❖ **Final Letter Grade:**
 - No pre-defined cutoffs. Will be based on bell curve of your performance.
 - Undergraduate and graduate students are compared in separate groups.

Blackboard



- ❖ Assignment instruction and files
- ❖ Submission (we don't accept email submission or hard-copy)
- ❖ Grades
- ❖ Questions, Discussion Forum

- ❖ Mobile App of BlackBoard

Deadlines



Everything will be submitted through Blackboard.

Due time: 11:59pm

Late submission: 5-point deduction per hour, till you get 0. 12:01am -5, 1:01am -10, ... (The raw score of each assignment is 100. So there is no point to submit it after 19 hours).

Regrading



7 days after we post scores on Blackboard. TA will handle regrade requests. Won't consider it after 7 days.

If not satisfied with the results, 7 days to request again. Instructor will handle it, and the decision is final.



Your Email

Make sure your UTA email account works. We will only contact you by your UTA email. Check it on a daily basis.



Schedule

<http://crystal.uta.edu/~cli/cse4334>

<http://crystal.uta.edu/~cli/cse5334>