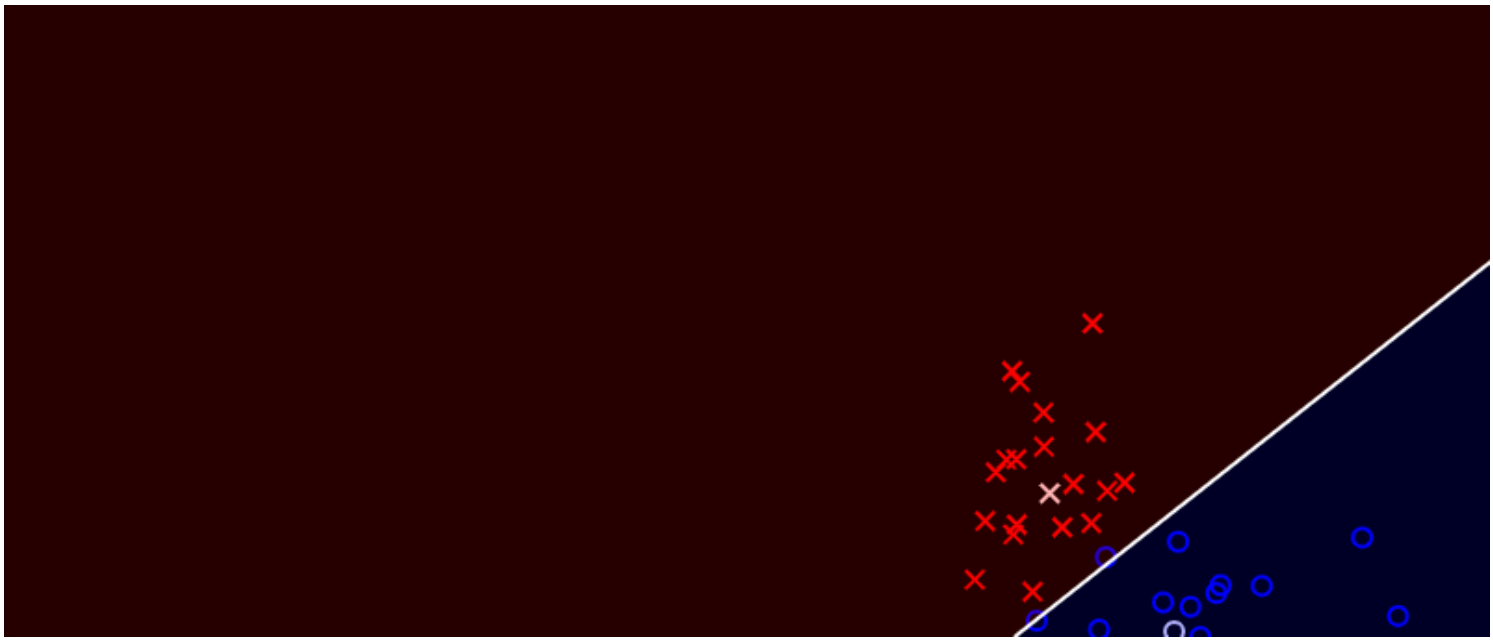

CS 504 Principles of Data Management and Mining — Fall 2020

Online. See https://cs.gmu.edu/~hrolenok/teaching/cs-504-fall2020/index.html for current syllabus. Be sure to check your @gmu.edu email, blackboard, and Piazza for updates.

Instructor: Brian Hrolenok
@gmu.edu email: hrolenok
Office Hours: 12:00pm-1:00pm, T/Th, ONLINE. See Blackboard for more information.

TA: Achange Destine
@gmu.edu email: adestine
Office Hours: 6-7pm M, 5-6pm F, ONLINE. See Blackboard for more information.

Note about online instruction

Due to the size of the class, and the limitations that places on the ability to maintain safe distance in the classroom, this course is being offered in an **online**, **asynchronous** format this semester. There are no scheduled lecture times, and all the material will be made available via Blackboard. Office hours will be held using videoconferencing software (tentatively Blackboard collaborate), and the instructor and TA will be monitoring Piazza and email for questions that arise outside of scheduled office hours.

Course description

CS 504 is a 3-credit introductory course that combines elements from a diverse field of topics, but with a slightly different focus than a course on databases or machine learning. In this course, the central object of study will be **data**: how to store and query efficiently, how to mine for hidden structure, and particularly the properties which make these tasks more or less difficult. Topics to include intro to database management, ER and relational data models, SQL and NoSQL, classification, and clustering. This course will include several individual programming and report based assignments, several quizzes, and a final exam. All assessments will be turned in online.

IMPORTANT NOTE: This course cannot be taken for credit by students of the MS CS, MS IS, MS ISA, MS SWE, CS PhD or IT PhD programs. Please check with the registrar if you are unsure about whether this course will count towards your degree program.

Learning objectives:

- To develop problem solving skills and analytical thinking from a data science perspective.
- To introduce a broad survey of modern approaches to data management and analysis.
- To develop the design and programming skills that will help you to interface with data at scale.

Prerequisites. The only prerequisite for this course is graduate standing, but please note that certain majors cannot take this course for degree credit (see catalog).

Textbook: There is no required textbook for this class. You may find the following text useful as a supplemental reference: Provost, Foster, and Tom Fawcett. Data Science for Business: What you need to know about data mining and data-analytic thinking. " O'Reilly Media, Inc.", 2013. ISBN:9781449361327

Homeworks

All assignment submissions will be handled through blackboard, and are due by the date and time listed there. Submissions by email will not be accepted.

Late Policy

You have three free late days to be used at your discretion thoughout the semester. That means you might turn in one assignment two days late or two different assignments one day late, etc. A free late day is "used" one minute after an assignment due date. A second free late day is "used" 24 hours and one minute after the due date. A third free late day is used 48 hours and one minute after the due date. After the free late days are exhausted, you will receive a 20% penalty per day.

Quizzes

Throughout the semester, there will be several *participation quizzes* given via blackboard. These will be short, multiple-choice, and you will receive full credit as long as you complete the quiz by the given due date.

The final exam will be given during the final exam period (Dec 9-Dec 16), and you will be able to pick a time within a window. There will be no make up for this exam unless previously arranged *(well in advance)*.

Grading policies

Your TA and I will strive to provide you reasonably detailed and timely feedback on every assessment. If you have any questions about any of your grades please reach out to us, either by coming to scheduled office hours or via your "@gmu.edu" email address. If there is an error with your grade, please contact us within a week of when feedback is returned, otherwise we might not be able to change it.

Point breakdown:

- Homeworks: 15% each (75% total)
- Participation quizzes: 10%
- Final exam: 15%

Academic Integrity

Please familiarize yourself with both the University wide honor code, as well as the one specific to the department of Computer Science. Violations of academic integrity will be reported to the Honor Committee. Course failure is a common recommended outcome for students found in violation.

Accommodations

If you need academic accommodations, please make sure you contact the instructor at the beginning of the semester or as soon as possible. Also make sure to contact GMU's Disability Services, available online (https://ds.gmu.edu, ods@gmu.edu) and by phone (703.993.2474), which coordinates all academic accommodations. After you have contacted ODS, you still need to contact the instructor so that appropriate arrangements can be made.

Topic outline

- Introduction to database management
- ER & EER model
- Relational data model and ER & EER to relational mapping
- SQL
- NoSQL
- Introduction to data mining concepts
- Classification
- Clustering