Data Mining Intro – Fall 2021

AW

Instructor and Tutor

- Prof. Alex Wolpert
  - **Primary:** Computer Science Theory + everything AI
  - **Secondary**: Linear Algebra, Statistics
  - Email: awolpert@roosevelt.edu
  - Web page: https://www.roosevelt.edu/academics/faculty/profile?ID=awolpert

Meetings and Attendance

- Class is Face to Face
  - This means you must be present, attendance taken
  - You can miss 3 sessions, no questions asked
  - Beyond 3 sessions the only reason to miss is documented sickness. If you need to miss for all other reasons you must obtain my permission no later than 1 day before.
  - Violators of this policy will incur stiff penalties, including getting F in the class for repeated violators
- We meet every Tue, Th 2:00pm-3:15pm
- Room AUD 524

Contacting the Instructor

- **The preferred method of communications with me during the semester is through email**
- You can also ask publicly accessible questions on Bb discussion board, but please notify me by email that you asked public question
- I guarantee email response within 24 hours during the week

Accessing Class Information

- Bb course site is the place to go. You can find there
  - Syllabus section that contains
    - Syllabus and schedule and HWs quantity,
    - All may change depending on class progress.
    - Textbooks in use
  - Weekly lectures contain lecture slides used in the lecture. Next week's lectures will be posted on the weekend after each week
  - Additional Materials section contains links to papers, websites and my handouts (slides and edited materials for independent study (especially useful to grad students). Area updated whenever need be. Check often.

Class Blackboard Site continued

- Bb course site is the place to go. In addition to syllabus, weekly lectures and additional materials sections, you can find there
  - Software section contains description of software in use and installation instructions. Almost no changes in this area throughout the semester,
  - Assignments section contains HWs. It is updated whenever new HW is posted (i.e. almost every week).

Office Hours

- You can come to see me Tue, 12:55pm-1:55 (**before** the class), but you need to make an appointment:
    - Please drop me an email that you are coming no later than on M, 7pm if you plan to see me on Tu.
- We can also meets any weekday on zoom ***if I have time***
    - If you intend to see me on a weekday then drop an email to me requesting an appointment 1 day before you'd like to see me, no later that 6pm. We'll negotiate the time for next day if I have time available. If not it'll be the following day
- You can ask questions by email

Home Assignments

- Homework assignments
  - Dates (available/due) already on the syllabus
  - Posted on Bb site in ‘Assignments’ content area.
  - Submission usually next week after posting on the Bb site with some exceptions due to Thanksgiving
  - Absolutely no late submissions accepted

Home Assignments

- Homework assignments:
  - Paper and pencil
    - 7 undergraduate/graduate;
      - Between 6 and 15 pts each – total approx. 70 pts;
    - 5 graduate only
      - Between 3 and 8 pts each – total 30 pts;
  - Programming undergraduate/graduate:
    - 5 between 4 and 8 pts – total 30pts.
- HWs will be posted on Bb site on Friday by 6pm

Grad Students Info

- Grad students will have to learn some material on their own. Note that they will be tested on this material
  - There will be clearly marked additional assignments that are for grad students only
  - There will be grad students only questions on exams
- Standards of grading are different for grad and undergrad students

Exams

- Exams:
  - 2 midterm exams - 70 pts each;
  - Final exam - 100 pts.
- !!!**Exams must be taken in-erson during the posted hours**!!! Dates are clearly specified in the syllabus.
- The only exception are for being sick or emergency. I will ask for documentation.
  - In this case we'll negotiate make-up time
- If you have a ***really serious reason*** because of which you can't take an exam in the scheduled time you must ask for my permission at least a day before the exam date.

Bonus Points

- Class participation is encouraged – you can earn bonus points
  - Solving a problem in class – 1,2 or 3 pts depending on the difficulty.
- Other bonus points
  - bonus problems on HW
- No other special assignments for extra credit will ever be given, especially not at the end of the term. No make up assignments will be given either
- The course is graded by accumulated points.

Grading Scale

- A ≥ 93% A- ≥ 89%
- B+≥ 87% B ≥ 82% B- ≥ 79%
- C+≥ 76% C ≥ 70% C- ≥ 65%
- D+≥ 55% D ≥ 50%

Prerequisite knowledge

- For having necessary skills:
  - Math 246 Linear Algebra AND CST 150 CS I AND

    one of the following
    - Math 217 Intro to Probability
    - Math 346 Probability and Stats I
    - Math 245 Discrete Math
    - Math 290 Into to Proof

Prerequisite knowledge

- By topic:
  - Linear Algebra:

    Distance/norm, orthogonality, best approximation for inconsistent systems of linear equations, diagonalization, eigenvalues/eigenvectors, inner products
  - Probability

    Counting probability, discrete probability distributions, expectation, variance/deviation, covariance, confidence interval.
  - Programming:

    Loops, functions, arrays

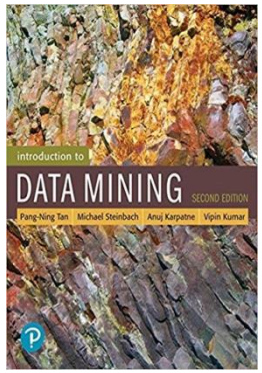
Things to Do Beyond HW

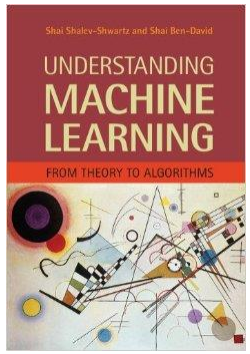
- You must check class Bb site at least twice a week
  - For new lectures, and HW (preferably Friday)
  - For discussions on discussion board (I suggest W)
- You are responsible for checking email and Bb class site frequently enough to stay current on all information pertaining to the course

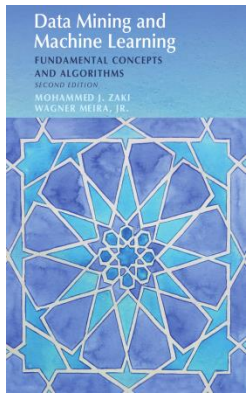
Textbooks

1. P.-N. Tan, M. Steinbach, Anuj Karpatne, V. Kumar. Introduction to Data Mining, Pearson, 2019
2. S. Shalev-Schwartz, S. Ben David. Understanding Machine Learning, 2015, CUP. Draft available on author's website free

   http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf
3. M. Zaki, W Meira. Data Mining and Analysis, Cambridge University Press, 2020
   https://dataminingbook.info/book_html/







Programming

- Yes, all students (not only CS and grads) must learn how to do programming in R
- The installation instructions for R, and its ecosystem can be found in 'Software' content area of the BB site. All links are provided
- You will also find links to tutorials for R
- We'll be learning it 'pattern style':
  - In class I'll provide examples of use, and references to examples on the web
  - You will learn how to program 'cut-and-paste' style by appropriate editing already know programming patterns