CS6220/DS5230 Unsupervised Data Mining, SUMMER 2022

About 6220/5230 Home Schedule Piazza Gradescope DM Resources

Instructor: Virgil Pavlu **(E-mail:** v.pavlu@northeastern.edu) but I prefer Team chat to email

Lecture: Mon / Wed , 1:30-3:10 (Boston Campus, WVG 102).

Office hours: Thu 6pm

TAs:

- Yashvi Bhandari, bhandari.ya@northeastern.edu
- Rohit Gupta, gupta.rohi@northeastern.edu
- Mingni Luo, luo.min@northeastern.edu

There will be about 5 helping sessions a week for students to discuss assignments and demo the code for grades. Students are encouraged to come in working-groups (if they work together on assignments).
SL room 009, 5pm

Discussion : piazza

Please use the discussion forum for all questions regarding material, assignments, due dates, data issues, programming issues, etc. That is, do not use the direct email to TAs or Instructors for these questions. Personal/private matters, such as availability, delays, grades, term projects or other advanced material, etc, should be discussed by direct email or team-chat.

MS Teams : Khoury - CS 6220 1 (Spring 2021)

Texts Recommended (not mandatory)

[A] Charu C. Aggarwal, Data Mining: The Textbook, Springer 2015

[LRU] Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman Mining of Massive Datasets, Cambridge University Press, 2014

Prerequisites

Linear algebra tutorial

http://www.stanford.edu/class/cs229/section/cs229-linalg.pdf

Probabilities tutorial

http://www.stanford.edu/class/cs229/section/cs229-prob.pdf

Programming: you should be familiar with rogramming in an imperative language like Java or Matlab or Python or R. While the programing language is your choice, we strongly encourage you to use one of these three mentioned. If you choose not to use on of these, discuss with the instructor first.

Homework

There will be seven assignments (one per module, every two weeks), including programming ; please see the syllabus for an exact schedule. Some of the problems can be difficult, and it will often be helpful to discuss them with others. Feel free to form study groups. However, the idea is for everyone to understand the problems and experience working through the solutions, so you *may not* simply "give" a solution to another classmate. In particular, each student *must* write up his or her own homework solutions/code and *must not* read or copy the solutions/code of others. If you work with others on a problem, you *must* note with whom you discussed the problem at the beginning of your solution write-up.

We expect the average total load to be 20 hours/week. Based on past experience, many students spend more than 20 hours/week for additional work denoted "Extra Credit"; the EC can improve your grade, but it is not a replacement for regular credit; EC is harder, more time consuming, worth less points, and less discussed during helping sessions; it is designed for students who want to get more out of the course. **The best way, by far, to maximize your grade is to complete the regular credit on time.**

Late homework policy: Homework is due at the *beginning of class* on the announced due date. You will be granted one homework extension of 1 week, to be used at your discretion, no questions asked. This policy does not apply to projects. After the first late assignment, unexcused late assignments will be penalized 20% per *calendar day* late. I normally will not accept assignments after the date on which the following assignment is due or after the solutions have been handed out, whichever comes first. If you will have a valid reason for turning in an assignment late, please see me *in advance* to obtain full-credit.

Projects

Optionally students can work on a project, if they discuss it in advance with the instructor .

Academic Honesty

All work submitted for credit must be your own.

You may discuss the homework problems or projects with your classmates, including approach, plans on paper/board, results, parameters, setup, analysis. **You cannot share code.**
You must acknowledge the people with whom you discussed your work, and you must write up your own solutions. Any written sources used (apart from the text) must also be acknowledged; however, you may *not* consult any solutions from previous years' assignments whether they are student or faculty generated.