CS 617: Data Mining

Dr. Truong Son Hy

Fall, 2023

Course Description

This course aims to equip students with basic knowledge and programming skills in Data Mining, Data Science, Machine Learning and Deep Learning for their further studies at graduate level, research or industry jobs (e.g., ML engineer, data scientist, etc.).

Location & Time

- **Time:** Monday / Wednesday, 11:00am - 12:15pm
- **Location:** Room A017, Root Hall
- **Online (via Zoom):** Video recordings will be provided.

`https://indstate-edu.zoom.us/j/94259887905`

Office hours

- **Time:** Tuesdays (weekly), 2:00pm - 5:00pm, First-Come-First-Serve (FCFS)
- **Location:** Room A181, Root Hall
- **Online (via Zoom):**

`https://indstate-edu.zoom.us/j/98501274920`

Programming Prerequisites

We will use Python programming language, Numpy, Sklearn, PyTorch.

Learning Outcomes

Upon successful completion of this course, students should be able to:

- Perform basic data cleaning operations, statistics and data visualization.
- Use tools in Python such as Numpy and Sklearn.
- Understand and use the basic regression, classification and clustering algorithms.
- Use basic data dimensionality reduction methods.
- Understand well-known Deep Learning methods to "mine" fundamental data forms such as images and texts.
- Utilize Deep Learning frameworks such as PyTorch.

Optional Textbooks

- *The Elements of Statistical Learning*, Trevor Hastie, Robert Tibshirani and Jerome Friedman, `https://hastie.su.domains/ElemStatLearn/`
- *Machine Learning: A probabilistic perspective*, Kevin Patrick Murphy, `https://probml.github.io/pml-book/book0.html`
- *Deep Learning*, Ian Goodfellow, Yoshua Bengio and Aaron Courville, `https://www.deeplearningbook.org/`

Expected Amount of Work

If you take this class seriously and get what you should out of it, some weeks you will likely be spending an average of 15 hours/week or more on the class. The students who get A's in their CS courses and have an easy time finding jobs spend at least this much time. Not everyone will need to spend this much time and not all weeks will be the same, but you should plan on putting in whatever time it takes. **Note:** your classes should be more important than your part-time job.

Course Annoucements

Announcements regarding the course will be made both during class, and on the Canvas page. You should regularly check these locations. Communications related to grades will be sent to your @sycamores.indstate.edu email address.

Course Outline

The main topics in the course are listed below. This is intended as a topical outline, not a timeline. Data Mining and Machine Learning concepts will be developed together.

Course Topics

- Introduction to useful tools in Data Science:
  - A review of mathematical concepts needed for the class: matrix operations, gradients, and elementary statistics.
  - A brief introduction to the Python programming language: basics, lists, dictionaries, comprehensions and functions/
  - Survey of important Python modules: Numpy, Pandas, Matplotlib.
- Basics of Machine Learning:
  - Linear regression
  - Gradient descent
  - Logistic regression
  - Neural networks (Multilayer Perceptrons)
  - Nearest neighbors
  - k-nearest neighbors
  - k-means clustering
  - Decision trees
- Deep Learning:
  - Convolutional Neural Networks
  - Recurrent Neural Networks
  - Image mining
  - Text mining
  - Knowledge (graph) mining

Assignments

The students in this course have the following responsibilities: read assigned readings before lecture, attend lecture, complete homework assignments (theory & practice), take the midterm, and take the final exam, and complete all projects. In this class, your final grade will be a weighted sum of all these components.

Policies

Note that this course follows all standard CS course policies. In particular check the CS course policies related to - cheating/plagiarism, attendance, missing exams. See `http://cs.indstate.edu/info/policies.html` for details.

All assignments are posted in a pdf file on the class web page. Each such file will indicate the number of points, the due date and time, and the location where your assignment should be saved. Failure to save your work in the correct location will be viewed as equivalent to not doing the work.

Grade Components

- Homework assignments (theory & practice): 30%
- Projects (programming, writing a report and presentation): 20%
- Midterm: 20%
- Final exam: 30%

Late Assignments

You have to submit your work by the given deadline. No extension.

Grading Policy

We try to design homework assignments and exams so that a standard cutoff for grades will be close to what you deserve. After the first exam a grade will be created in Blackboard called **Letter Grade** that is intend to be your current grade in the class. The grades are generally based on the following table.

Letter Grade	**Percentage**
A	93-100
A-	90-93
B+	87-90
B	83-87
B-	80-83
C+	77-80
C	73-77
C-	70-73
D+	67-70
D	63-67
D-	60-63
F (Fail)	0-60

Grades are intended to indicate your mastery of the course material. The following are offered as guidelines.

A
The student can do all the assignments on your own.

A-/B+
The student understands almost everything, and should be able to use this knowledge in other courses or in a job.

B/B-
The student understands most, but not all, topics well.

C+/C
The student has learned enough and got the minimum skills to move on in the subject.

C-/D+
The student made some effort in, and understands some things at a high level, but hasn't mastered the details well enough to be able to use this knowledge in the future.

D/D-
Students will normally not get an F if - they attend 80% of the lectures, complete most of the assignments up through the end of the course, and get around half of the problems on the midterm and the final exam correct.

F
Normally, students that get an F simply stopped doing the required work at some point.

Academic Integrity

Read CS course policies in terms of what is and is not allowed on assignments: `http://cs.indstate.edu/info/policies.html`. Please ask the instructor if you have doubts about what is considered cheating in this course.

Special Needs / Student Disabilities

Indiana State University recognizes that students with disabilities may have special needs that must be met to give them equal access to college programs and facilities. If you need course adaptations or accommodations because of a disability, please contact us as soon as possible in a confidential setting either after class or in my office. All conversations regarding your disability will be

kept in strict confidence. Indiana State University's Student Support Services (SSS) office coordinates services for students with disabilities: documentation of a disability needs to be on file in that office before any accommodations can be provided. Student Support Services is located on the lower level of Normal Hall in the Center for Student Success and can be contacted at 812-237-2700, or you can visit the ISU website under A-Z, Disability Student Services and submit a Contact Form. Appointments to discuss accommodations with SSS staff members are encouraged. Once a faculty member is notified by Student Support Services that a student is qualified to receive academic accommodations, a faculty member is obligated to provide or allow a reasonable classroom accommodation under ADA.

Disclosures Regarding Sexual Misconduct

Indiana State University fosters a campus free of sexual misconduct including sexual harassment, sexual violence, intimate partner violence, and stalking and/or any form of sex or gender discrimination. If you disclose a potential violation of the sexual misconduct policy I will need to notify the Title IX Coordinator. Students who have experienced sexual misconduct are encouraged to contact confidential resources listed below. To make a report or the Title IX Coordinator, visit the Equal Opportunity and Title IX website: `http://www.indstate.edu/equalopportunity-titleix/titleix`.

The ISU Student Counseling Center

HMSU 7th Floor, 812-237-3939, `www.indstate.edu/cns`.

The ISU Victim Advocate

- Leah Reynolds `leah.reynolds@indstate.edu`, HMSU Room 813, 812-237-3829 (office), 812-243-7272 (cell).
- Trista Gibbons `trista.gibbons@indstate.edu`. HMSU 7th Floor, 812-237-3939 (office), 812-230-3803 (cell).

United Campus Ministries

321 N 7th St., Terre Haute, IN 47807 812-232-0186

`http://www2.indstate.edu/sao/campusministries.htm`
`http://www.unitedcampusministries.org`