CS 457: Machine Learning

G. Exoo

Spring, 2022

E-mail: cs457@cs.indstate.edu   Web: **cs.indstate.edu/cs457**
Office Hours: TBA   Class Hours: MWF 11:00-11:50AM
Office: A-179 Root Hall   Class Room: Zoom Meetings
Final Exam: May 6, 10:00am.

Course Description

CS457 provides an introduction to databases and data models. It will introduce students to the foundations of data base systems and cover basic topics such as relational algebra, the evolution of data models, and techniques of query optimization and processing. Students will have an opportunity to develop their skills using technologies such as SQLITE, Postgres, MySQL, MongoDB and Redis.

The principal goals will be:

- Learn to effectively use SQL to create, query and maintain databases.

- Design relational databases.

- Learn the basic alogorithms used in database implementations.

- Understand the characteristics of various classes of databases, including NoSQL databases.

- Learn the JSON format and how to query Mongo databases.

- Learn how to use Redis and understand some of the basic algorithms available in Redis.

Course Outline

- Week 1 - Class protocols, course objectives, using Linux.

    - Summary
        * We will discuss the database concept, what capabilities we require in a data base, and a few different types of databases.
        * There will be a problem set on relational algebra.
    - Overview of Data Models and Databases

- ∗ Relational Model - our Topic for the first half of the course
  - · Relations
  - · Relational Algebra
  - · Examples of Relational Databases: SQLite, Postgres, DB2, Oracle, Sybase, MySQL, MariaDB
- ∗ Other Database Models
  - · Hierarchical Model
  - · Network Model
  - · Semantic Model
  - · Document Model
  - · Object Oriented Model
  - · Semi-structured Models

- Week 2 - SQLite: the Basics

  - Summary

    - ∗ We will learn to use SQLite3, and begin learning SQL.
    - ∗ There will be a programming assignment that involves reading a data file into an SQLite table and doing some SELECT queries.

  - What is SQL?

    - ∗ SQL is each of the following: DML, DDL, DCL, DQL, TCL

  - What is SQLite?

  - How to Use SQLite

    - ∗ Databases and Tables
      - · Creating tables
      - · Importing data

  - Basic queries

    - ∗ CREATE, INSERT, SELECT

  - Modifying tables

    - ∗ INSERT, UPDATE, DELETE

- Week 3 - Complex Queries

  - Summary

    - ∗ We will finish a first pass over the SQL language.
    - ∗ There will be a programming assignment involving multiple tables.

  - The SQL Language

    - ∗ Data Definition Language: CREATE, DROP
    - ∗ Data Query Language: SELECT
    - ∗ Data Manipulation Language: DELETE, INSERT, UPDATE
    - ∗ Data Control Language: GRANT, REVOKE

         * Transaction Control Language: ROLLBACK, COMMIT, SAVEPOINT
- – SQLite Syntax Diagrams

- Week 4 - SQL Practice and Triggers

  - – Summary

    * Triggers will be introduced. Triggers are one way to turn a static database into a dynamic entity.
    * The week 3 assignment will be expanded using triggers.

  - – Triggers

    * Functions that are automatically executed in response to a database event.
    * DML triggers.
    * DDL triggers.
    * Other events.

- Week 5 - Efficiency: Data Base Design

  - – Summary

    * Design databases that maximize integrity and minimize redundancy
    * Begin the database design project.

  - – Define the goals for the database design.

    * Online Transaction Processing
    * Decision Support Systems

  - – Normalization

    * Example of Normalization Process
    * Anomalies: Insertion, Deletion, Update
    * Extensibility

- Week 6 - Building a Working Data Base

  - – Summary

    * Apply Normalization Process to our design.
    * Complete the database design process

- Week 7 - Database Internals

  - – Summary

    * Study some of the algorithms and data structures used in databases.
    * The project for the week will involve implementing a tree data structure in C.

  - – Algorithms and Trees

    * B Trees
    * B+ Trees
    * Tries

  * Hash tables
  * Hash buckets
  * Heap files

- Week 9 - Relational Client Server Databases

  – Summary

    * Study the new issues related to the client/server model.
    * Implement an earlier project in a client/server environment.

  – Brief discussion of MYSQL, MariaDB, Oracle, Postgres, etc.
  – Concurrency issues.

    * Consistency.
    * Lost data.

  – Examples and Exercises MariaDB or Postgres.

- Week 10 - Data Exchange Formats - XML and JSON

  – Summary

    * A brief review of Javascriptm, followed by a study of JSON Data structures and files.
    * The assignment will involve parsing JSON in a high level language.

  – Using JSON

    * Javascript Examples
    * Python Examples

- Week 11 - NoSQL: Mongo

  – Summary

    * Introduction to the MongoDB and the Mongo Shell
    * The project for the week will involve doing a list of query exercises on a database created by the instructor.

  – Mongo, JSON and Javascript

    * The Mongo Shell
    * MongoDB CRUD operations.
    * Sharding and an introduction to big data.

- Weeks 12 and 13 - Mastering Mongo Syntax

  – Summary

    * Learn to use the aggregation pipeline and the Map-Reduce perspective.
    * The project for the week will involve complex, multi-stage queries.

  – MongoDB

    * Map-Reduce

* Aggregation pipeline
* Operations

- Week 14 and 15 - More on Big Data

  – Summary

    * A brief introduction to Data Mining.
    * There will be a programming assignment on a big data problem.

  – The Stream Data Model

    * Sub-log-space counting.
    * Exotic hashing methods.
    * Detecting duplicates.
    * Heavy hitters.

Recommended Textbooks

There is no textbook. Online readings will be suggested for each topic covered.

Online Texts/References

These may be useful for topics that are completely new to you.

The Standard Python Tutorial

Virtual Environments

Introduction to Numpy

Linear Algebra in Numpy

Matplotlib Documentation

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 Announcements

Announcements regarding the course will be made both during class, on the class web page, and to your cs email (which you should learn how to use). Communications related to grades will be sent to your @sycamores.indstate.edu email address. You should regularly check these locations.

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.

Grading

Grade Components

Quizzes - 20

Programming Projects - 60

Final Exam - 20

The students in this course have the following responsibilities: read assigned readings before lecture, attend lecture, complete homework assignments, take in-class quizzes, take exams, and complete all projects. In this class, your final grade will be based primarily on the projects, and to a lesser extent on the exams and quizzes.

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.

Late Assignments

The maximum points you will receive for late assignments will decay exponentially with time. If n is the number of points the assignment is worth, then a perfect assignment that is between $d - 1$ and d days late will net at most $n/(2^d)$ points. [1]

We suggest attempting a homework assignment the day it is given, or the day after, so that if you do not understand how to do the assignment, you will have time to seek help. You may need to ask for help more than once, and you should certainly plan on spending a lot of time in the CS lab (A-015 Root Hall). Many of the homework assignments require thought and creative problem solving. Do not expect to solve the problems the first time you attempt them.

---

[1]Quiz: If an assignment is worth 100 points and you hand it in one week after it is due, what is the maximum number of points you could receive?

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.

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	0-60

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

A

You can do all the assignments on your own.

B+/A-

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+

Learned enough and have the minimum skills to move on in the subject.

D+/C-

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-

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

F

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

Special Needs / Disability Services

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.

COVID-19 Information

Students are expected to adhere to course attendance policies, as stated in the course syllabus. Students must complete the Sycamore Symptom Assessment by email before arriving on campus each day unless they have documented their COVID immunization and have been exempted from the program. Documented COVID-related absences will be treated like any other serious medical issue. Following University policy, students with a documented, serious medical issue must contact the Office of the Dean of Students for assistance. The Office of the Dean of Students will supply documentation for faculty. Students with a documented serious medical issue should not be penalized and will be given a reasonable chance to complete exams or assignments. Once notification is made, faculty will make reasonable efforts to accommodate the student's absence and will communicate that accommodation directly to the student. Please note that faculty are not required to accommodate a serious medical issue with virtual content options, like streaming or recorded lectures. Students who have been notified by contact tracers of a close contact who has tested positive for COVID must abide by their instructions, which will include a mandatory period of quarantine, especially if the student is unvaccinated, and/or mandatory testing. To avoid the potential of missing significant class time, students are strongly encouraged to receive the COVID vaccination that has been made available on campus. For more information about

the vaccines or to find a vaccination site, go to: http://ourshot.in.gov. The ISU Health Center also administers COVID-19 vaccines by appointment.

Students should contact the Office of the Dean of Students with questions by calling 812-237-3829.

The information provided in this section of the syllabus is subject to modification based on guidance by public health authorities. Please follow this link for updated information on ISU's Fall 2021 requirements.

Masks on campus: All faculty, staff, and students are required to wear face coverings anytime they are in public spaces. In classrooms they are required of faculty, staff, and students. Failure to comply with this policy will be treated, by policy, as would any student disruption of class. (A refusal or failure to properly wear an appropriate mask will result in the faculty member asking the student to do so. A refusal of that request may result in the student being asked to leave the class for that period. A refusal of that request may result in the faculty member cancelling class and referring the matter to the Office of Student Conduct and Integrity.) A failure of a faculty member to wear an appropriate mask/shield should be reported to the Chairperson of the faculty member.

Face coverings will be worn by all students and faculty in classrooms as well as in buildings (unless you are alone in an office). What is said/printed on a mask will be held to the same Student Code of Conduct standard as if it were printed on a shirt or hat. As a result, a political statement such as MAGA, BIDEN2020, or BLM is not grounds for demanding that it be removed/replaced. In judging what constitutes an offensive statement on a mask, the determination will be made by Student Affairs using the Student Code of Conduct. If there is a question about a mask, the faculty member will refer the matter to Student Affairs and only insist upon its immediate removal if there is no doubt that it violates the Code. Medical waivers will be made through Student Affairs and students with such a waiver are expected to carry the documentation with them and present it when asked.

Academic Integrity

Please follow these guidelines to avoid problems with academic misconduct in this course: Homework: You may discuss the homework assignments, but should solve and finish them on your own. To make sure you are not violating this, if you discuss with someone, you should DESTROY any work or evidence of the discussion, go your separate ways, SPEND at least an hour doing something completely unrelated to the assignment, and then you should be able to RECREATE the program/solution on your own, then turn that in. If you cannot recreate the solution on your own, then it is not your work, and you should not turn it in. Note on sources: if you use some other source, the web or whatever, you better cite it! Not doing so is plagiarism. Exams: This should be clear no cheating during exams. The exams will be closed-book, closed-notes, no computer, and no calculator. You may be allowed one sheet of 8.5" by 11" piece of paper with hand-written notes to use as a crib-sheet for your tests. Projects: You should not copy from the Internet or anywhere else. The project should be your own work. It will be fairly obvious to me if you do copy code from the Internet, and the consequences will be at the least a 0 on the project. If cheating is observed, you will at the least receive a 0 for the assignment (and may receive an F for the course), and I will file a Notification of Academic Integrity Violation Report with Student Judicial Programs, as required by the university's policy on Academic Integrity. A

student who is caught cheating twice (whether in a single course or different courses) is likely to be brought before the All University Court hearing panel, which can impose sanctions up to and including suspension/expulsion. See the http://www.indstate.edu/sjp/docs/code.pdf and http://www.indstate.edu/academicintegrity/ for more information.

Please ask the instructor if you have doubts about what is considered cheating in this course.

Statement on Discrimination, Sexual Harassment, and Sexual Misconduct

Indiana State University Policy 923 strictly prohibits discrimination on the basis of: age, disability, genetic information, national origin, pregnancy, race/color, religion, sex, gender identity or expression, sexual orientation, veteran status, or any other class protected by federal and state statutes in ISU programs and activities or that interferes with the educational or workplace environment.

Title IX of the Educational Amendments of 1972 prohibits discrimination based on sex, including sexual harassment. Sexual harassment includes quid pro quo harassment, unwelcome verbal or physical conduct, sexual assault, dating violence, domestic violence, and stalking.

If you witness or experience any forms of the above discrimination, you may report to:

- Office: Equal Opportunity & Title IX; (812) 237-8954; Rankin Hall, Room 426

- Email: mailto:ISU-equalopportunity-titleix@mail.indstate.edu

- Online: https://cm.maxient.com/reportingform.php?IndianaStateUniv&layout_id=10

Disclosures made to the following confidential campus resources will not be reported to the Office of Equal Opportunity and Title IX:

- ISU Student Counseling Center: (812) 237-3939; Gillum Hall, 2nd Floor

- Victim Advocate: (812) 237-3829; HMSU 7th Floor

- UAP Clinic/ISU Health Center: (812) 237-3883; 567 N. 5th Street