

Syllabus

Instructor

Dr. K. Kalpakis
Computer Science & Electrical Engineering Department
Office: ITE 348
Phone: (410) 455-3143
Email: kalpakis@umbc.edu
Class homepage: <http://www.csee.umbc.edu/~kalpakis/Courses/491-fa17/>
Office Hours: Tuesday and Thursday, 1:15pm-2:15pm, and by appointment.

Meeting Time and Place

Tuesday and Thursday 10:00am-11:15am
Room Physics 101

TA: Vedmurtty Chavan

Room: ITE 344E, Email: ved1@umbc.edu
Office hours: Tuesday and Thursday 11:30am-12:30pm, and by appointment.

Important Dates

- Last class, December 12.
- Final Exam, Tuesday, December 19, 10:30am-12:30pm.

Prerequisites CMSC-341, MATH-221, and STAT-355 or permission of the instructor.

Text. The recommended textbooks are listed on the course website.

List of Topics (tentative).

- Data Science process
- SQL and NoSQL databases
- Python review, Pandas
- Data types, basic statistical descriptions of data, data similarity metrics
- Data visualization
- Data preprocessing (cleaning, integration, reduction, transformation, discretization)
- Model building, evaluation, cross-validation, and selection

- Clustering
- Linear regression
- Classification
- Outlier detection
- Active learning, Transfer learning, and Deep Learning
- Dimensionality reduction
- Recommendation systems
- Graph Mining
- Scaling techniques
- Map-Reduce, Hadoop, and Spark
- Special topics (TBA)

Course objectives:

The goal of this class is to give students an introduction to all the phases of the data science process using real data and modern tools. Topics that will be covered include data formats, loading, and cleaning; data storage in relational and non-relational stores; data analysis using supervised and unsupervised learning, and sound evaluation methods; data visualization; and scaling up with cluster computing, MapReduce, Hadoop, and Spark.

Required Work

Required work consists of (1) final exam, (2) homework assignments, and (3) a project. Further, you are expected to actively participate in class discussions.

Students registered for graduate credit (CMSC 691), will have modified assignments (homeworks, exams), and have higher expectations from their project along all dimensions (innovation, difficulty, soundness, completeness, design, analysis, implementation, evaluation, and presentation).

Homework assignments will help you master the fundamental concepts of database systems, while the project assignment will help you in developing skills to apply these concepts in practical data management applications. The exams will be yet another opportunity to receive feedback with respect to your accomplishment of the course objectives.

Assignments

There will be at least four homework assignments. Homework assignments involve design, analysis, and software development and will need to be completed in a short period of time, usually in about a week.

There will also be a semester long project assignment, which will have a design, analysis, and implementation component. Basic programming skills using the Python programming language, together with various python libraries, are needed in the project.

Projects will be done in teams of 3-5 students, enabling the students to practice and demonstrate the data science process with real data on an application of their choice. Each team will present its project to the whole class during the last two class periods of the semester. Additional details about the projects will be provided during the first few weeks of the semester.

Ground Rules for Assignments

- Assignment details, due dates, etc will be posted at the class homepage. Students are strongly advised to check the class homepage on a regular basis. Failure to do so is not an acceptable excuse for missing an assignment or for not adhering to the assignment's instructions.
- You may develop the programs, if any, for your assignments using the computers available to you at UMBC, or any other computer available to you. However, no matter what computer you use to develop your programs, you must make sure that your programs can run successfully on Virtual Machine (VirtualBox) running the latest Ubuntu Linux operating system.
- All assignments must be submitted electronically by the date they are due according to the assignment's submission instructions. No late assignments will be accepted, unless University Policy states otherwise.
- In submitting an assignment, students must adhere to the submission instructions specified by that assignment.
- The written part of each assignment must be typed using a word-processor of your choice (you may include hand-written mathematical formulas and/or diagrams as images in your documents). No matter how you prepare the written part of your assignment, it must be submitted in the Adobe PDF format. No other formats are going to be accepted.
- No collaboration. Unless otherwise specified, each assignment is to be done and written individually by each student. Students should not collaborate on any assignment. The only exception would be for assignments for explicitly designated as team assignments, where team members are expected to collaborate in completing such an assignment.
- Students may be asked to come in and explain their solution(s) to an assignment to the instructor(s) and/or TA(s). Failure to satisfactorily demonstrate authorship of a solution is a violation of Academic Integrity policy.

Students are strongly advised to keep up with the assignments and other coursework. Homework and project assignments do demand the amount of time allocated to them.

Exams

There will be one comprehensive final exam. All the exams will take place in class and will be closed-book and closed-notes. Make-up exams are possible only under University Policy. You should make prior arrangements with the instructor if you expect to miss an exam.

Each student should have his student photo identification card or driver's license when taking an exam. Failure to produce a proper photo ID may result in getting a zero on that exam.

Communication

Students are strongly advised to check the class homepage, and the course Blackboard area

| Activity | Weight |
|------------|--------|
| Homework | 20% |
| Project | 50% |
| Final Exam | 30% |
| Total | 100% |

Table 1: Course Activities and their relative weights.

<http://blackboard.umbc.edu>

on a regular basis for news, announcements, and assignments. Failure to do so is not an acceptable excuse for missing an assignment or announcement.

Students are welcome to use the course Blackboard area to discuss topic matters. However, student's are advised not to solicit or post solutions to any assignment or otherwise violate Academic Integrity policy.

Grading Policy

The course grades will be determined as follows. For each course activity in Table 1, each student will receive an activity score, which will be the average of the student's scores on the assignments for that activity. An activity score is a number in the range $0 \dots 100$. A term score will be computed by taking the weighted sum of the activity scores, using the relative weights given in Table 1. The instructor will convert term scores into letter grades by using the following mapping:

$$[90, 100] \Rightarrow A, [80, 90) \Rightarrow B, [70, 80) \Rightarrow C, [60, 70) \Rightarrow D, [60, 100] \Rightarrow P, [0, 60) \Rightarrow F.$$

Incomplete grades will issued only under those extreme situations described by University Policy for granting incompletes. Failure to complete assignments on time is not a sufficient reason for an incomplete.

Necessary, but not sufficient, conditions to pass the course are as follows: you must have a homework activity score of at least 50% of the points, at least 60% of the points for the project, and at least 30% of the points of the final exam.

Academic Integrity Policy

By enrolling in this course, each student assumes the responsibilities of an active participant in UMBC's scholarly community in which everyone's academic work and behavior are held to the highest standards of honesty. Cheating, fabrication, plagiarism, and helping others to commit these acts are all forms of academic dishonesty, and they are wrong. Academic misconduct could result in disciplinary action that may include, **but is not limited to**, suspension or dismissal. To read the full Student Academic Conduct Policy, consult the UMBC Student Handbook, the Faculty Handbook, or the UMBC Policies section of the UMBC Directory.

There is no tolerance for academic dishonesty in this course. Any and all academic dishonesty acts will be treated severely, as prescribed in the UMBC's Student Academic Conduct Policy.

ADA Compliance

We recognize that some of you may have disabilities that require special attention from the instructional staff. Please make us aware of them at your earliest so that UMBC can make suitable arrangements.