AMAT 502—Modern Computing for Mathematicians—Fall 2022 Syllabus

Instructor: Prof. Justin M. Curry   **Office:** CK 391   **Email:** jmcurry@albany.edu

Course Objectives: *Modern Computing for Mathematicians (AMAT 502)* is designed to take a student with no previous background in computer science or programming to a basic level of competency in programming in Python so that they can carry out the fundamental tasks of a data scientist. Specifically, by the end of the semester the student should be able to: define functions, assign values to variables, work with basic data types such as lists and dictionaries, understand concepts from object oriented programming (OOP) such as inheritance and polymorphism, create a class from as well as subclasses, call methods on objects, control the flow of a program via good programming principles such as abstraction, create scatter plots and plots of mathematical functions using Matplotlib, simulate random phenomena such as flipping a coin, apply machine learning methods such as regression, KNN, K-Means, PCA, logistic regression, SVM, and naive Bayes classifiers. The student will be able to load a data set (provided as a CSV file) as a Pandas data frame and perform basic data cleaning, feature engineering, and apply scikit-learn libraries to analyze said data set. A final project, which consists of a 20 minute group oral presentation and a 4-8 page writeup is meant to develop presentation skills using some slideshow software, e.g. PowerPoint or Key Note, as well written communication skills in English and basic competency in the LaTeX typesetting language. Working in groups provides opportunities for students to develop their interpersonal skills and prepare them for the workplace.

Description of the Course: *Modern Computing for Mathematicians (AMAT 502)* is meant to be the core requirement of the Data Science Masters program here at UAlbany. This course will provide an introduction to programming for students who have never coded before as well as a refresher for more experienced programmers. We will also provide a theoretical and practical introduction into machine learning. At a high-level, this course has three parts:

- 8 lectures on Programming Fundamentals in Python
- 4 lectures on Numpy, SciPy and Statistics
- 11 lectures on the machine learning using Pandas and Scikit-Learn as well as its conceptual underpinnings.

The remaining time is for group projects and presentations. A lecture outline is on Page 3.

Pre-Requisites: Basic undergraduate mathematics such as calculus, linear algebra, probability and stats.

Lecture Place and Times: 6716: *ES 232* on *MW 11:40-1pm*, 7412: *PH 224* on *MW 3-4:20pm*. You will need to download the Global Protect VPN (first link below) to access the appropriate server for your class when you're off campus. Your login is your netID **all lower case** and your usual password.

- https://wiki.albany.edu/pages/releaseview.action?pageId=77139467
- https://amat502-6716-fa22.its.albany.edu
- https://amat502-7412-fa22.its.albany.edu
- Backup Option: https://github.com/jmc42/AMAT502

Office Hours: Thurs 1-3pm in CK 391 or by appt with Robby Green rgreen@albany.edu or Tung Lam tlam@albany.edu. Students are encouraged to use the Discord Server. https://discord.gg/REc86JsN

Grading Schema:

- 30% Determined by problem sets, a programming project, and a proposal; denoted ♣.
- 30% Determined by 3 in-class open notes exams; denoted by ♢.
- 30% Determined by a group (3-5 people) final project; denoted by ♠.
- 10% Determined by attendance and participation. Absences must be approved by Dr. Curry.

Academic Integrity: Please familiarize yourself with UAlbany's *Standards of Academic Integrity* page: https://www.albany.edu/studentconduct/standards_of_academic_integrity.php

I take violations of academic integrity very seriously. Programming is a subject where it is too easy to cheat yourself of valuable learning opportunities. It is permissible to consult existing questions and answers on *Stack Overflow* or *Geeks for Geeks* or similar programming help websites, but you CANNOT post one of the questions from the HW or projects as a new question to one of these sites or use Chegg or a similar "homework help" service. You are, however, permitted to ask your classmates or me or the TAs for help, but you should make sure you've tried to figure it out on your own for 10-15 minutes at least.

The single most important thing to NOT do is to copy large blocks of code from a classmate or the internet. If you are caught copying and pasting someone else's code the following penalty system will apply:

(1) For the first offense, **a zero for the question that you copied on.**
(2) For the second offense, **a zero for the assignment and I file a Violation of Academic Integrity Report (VAIR).**
(3) For the third or later offense, **you get a letter grade reduction (or possibly fail) and I refer you to community standards, which could result in expulsion from UAlbany.**

Course Materials: Jupyter Notebooks cover all the necessary material for this class as well as videos for each lecture. Please watch these ahead of time.

- GitHub Back-Up of Lectures: https://github.com/jmc42/AMAT502/
- YouTube Lectures: https://www.youtube.com/channel/UC3HHUGPjUfyH0YmkYXfFMbQ
- Required Text: *Introduction to Computation and Programming Using Python* by Guttag, 3rd ed.
- Additional Free Text: https://github.com/jakevdp/PythonDataScienceHandbook

Attendance and Missed Work Policy:

Attendance during this course will be mandatory and counts for 10% of your grade. Valid excuses for illness and emergencies will be accepted, but must be reported before class or as soon as reasonably possible. It is considered the student's responsibility to communicate reasons for missed work and to follow university policy, but I will consider excuses on a case-by-case basis. See this this for official UAlbany policy.

https://www.albany.edu/graduatebulletin/requirements_degree.htm#attendance

There will be approximately five problem sets and an extended programming project. These range in value between 14 to 30 points. **I will take off 2 points off for every day an assignment is late, unless I grant you an extension, which must be negotiated before the due date.**

Lecture Outline:

The lectures are pre-recorded and available on YouTube. Search for "AMAT502". The following is the link to the channel. Please Like and Subscribe!

`https://www.youtube.com/channel/UC3HHUGPjUfyH0YmkYXfFMbQ`

- Lec 1 = Introduction to Data Science + Basic Numerical Operations,
- Lec 2 = Conditionals and Loops,
- Lec 3 = Bisection Search and Functions, Quick Discussion of Big-O Notation
- Lec 4 = Functions and Recursion,
- Lec 5 = More Recursion, Strings and Lists
- Lec 6 = Data Types and Edit Distances
- Lec 7 = "One Liners" and Intro to Object Oriented Programming
- Lec 8 = More OOP and Data Structures
- Lec 9 = OOP for Mortgages and Matplotlib
- Lec 10 = NumPy Array Operations + Intro to Randomness
- Lec 11 = Basic Probability Distributions Review
- Lec 12 = Hypothesis Testing and the CLT
- Lec 13 = Into to Machine Learning: Regression
- Lec 14 = Intro to ML: Classification and Clustering Overview,
- Lec 15 = Catch-Up Lecture + Intro to Pandas
- Lec 16 = K-Means Clustering
- Lec 17 = Principal Component Analysis
- Lec 18 = MNIST via K-means and K-Nearest Neighbors
- Lec 19 = Model Validation and Bias-Variance Tradeoff
- Lec 20 = Naive Bayes Classifier
- Lec 21 = Maximum Likelihood and Logistic Regression
- Lec 22 = SVM: From Hyperplanes to Kernels
- Lec 23 = Decision Trees and Entropy

Look above for the approximate content of the lectures listed below.

Monday		Wednesday	
Aug 22nd	Lecture 1	**Aug 24th**	Lecture 2
Aug 29th	Lecture 3	**Aug 31st** ♣**PS01 Due Friday 9/2**♣	Lecture 4
Sep 5th ♡*LABOR DAY = NO CLASS*♡		**Sep 7th**	Lecture 5
Sep 12th	Lecture 6	**Sep 14th** ♣**PS02 Due Friday 9/16**♣	Lecture 7
Sep 19th	Lecture 8	**Sep 21st** ◇**IN-CLASS EXAM 1: Lecs 1–6**◇	
Sep 26th	Lecture 9	**Sep 28th** ♣**PS03 Due Friday 9/30**♣	Lecture 10
Oct 3rd	Lecture 11	**Oct 5th**	Lecture 12
Oct 10th ♡*FALL BREAK = NO CLASS*♡		**Oct 12th** ♣**PS04 Due Friday 10/14**♣	Lecture 13
Oct 17th	Lecture 14	**Oct 19th** ◇**IN-CLASS EXAM 2: Lecs 7–12**◇	
Oct 24th	Lecture 15	**Oct 26th** ♣**Optional PS05 Due 12/12**♣	Lecture 16
Oct 31st	Lecture 17	**Nov 2nd**	Lecture 18
Nov 7th ♣**Project K-Means Due Monday 11/7**♣	Lecture 19	**Nov 9th**	Lecture 20
Nov 14th ♣**Final Project Proposal Due**♣	Lecture 21	**Nov 16th** ◇**IN-CLASS EXAM 3: Lecs 13–19**◇	Lecture 22
Nov 21st	Lecture 23	**Nov 23rd** ♡*THANKSGIVING = NO CLASS*♡	
Nov 28th *Presentation Prep*		**Nov 30th** ♠*Final Presentations Day 1*♠	
Dec 5th ♠*Final Presentations Day 2*♠		**Dec 7th** ♡*NO MORE CLASSES*♡	

♣ Final writeups and Optional PS05 are due no later than 11:59pm on Monday, December 12th, 2022. ♣