MAT 345 Syllabus

Semester: Fall 2018
Course title: Introduction to Data Science
Course Web Page: Moodle page for MAT 345
Time/Place: MW 9:00 - 10:20am in Von Newmann

Instructor Contact:
Dr. Brigitta Vermesi
Email: bvermesi@digipen.edu
Phone: (425) 629-5092
Office hours: MW 10:30-11:59AM, Th 12:00-1:00PM

Prerequisites:
MAT 140 and MAT 258

Course Description:
This course presents a variety of computational tools for modeling and understanding complex data. Topics include manipulating data, exploratory data analysis, statistical inference, spam filters and naïve Bayes, neural networks, and machine learning algorithms such as linear regression, k-nearest neighbors, and k-means. The course will focus on both understanding the mathematics underlying the computational methods and gaining hands-on experience in the application of these techniques to real datasets.

Course Objectives and Learning Outcomes:

Since the goal of this course is to ultimately be able to analyze large sets of data, part of the coursework will involve working in groups or individually to analyze data sets. To this end, students will review topics from linear algebra, probability and statistics and understand how they apply to the study of large data. Students will learn a variety of algorithms used in data science, code them and test them on real data sets.

Textbooks:
No textbook required.

Resources:
Doing data science, by R. Schutt, C. O'Neil
Data science from scratch, by Joel Grus
Learning from data, by Y. Abu-Mostafa, M. Malik-Ismail, H.T. Lin
The elements of statistical learning, by T. Hastie, R. Tibshirani and J. Friedman

Grading:

Midterm Exam	25%
Final Exam	35%
Homework	15%
Projects	25%

Grades will be determined based on total course percentage. Percentage scores will determine letter grades according to the scale: (in the worst case)

A 93 − 100
A- 90 − 92.9
B+ 87 − 89.9
B 83 − 86.9
B- 80 − 82.9
C+ 77 − 79.9
C 73 − 76.9
C- 70 − 72.9
D 60 − 69.9
F < 60

Exams: There will be a midterm exam given during regular class hours and a *comprehensive* final exam. There are NO make up exams unless you have a *compelling and well-documented reason* for missing a test.

Homework assignments: Homework will be assigned and posted on the web page. You are responsible for checking the web page and noting the assignments and the due date, as well as the submission format. You may work on homework together, as well as consult the tutors and the instructor. However, the final work that you turn in must be your own work.

Projects: There will be several projects in the course, illustrating various algorithms and methods applied to real data sets. They will be announced in class and posted on Moodle. Students will submit their assignments online, on Moodle. No late submissions will be accepted.

Technology in the Classroom:
The use of laptops, cell phones, smart phones, or other mobile communication devices is disruptive, and is therefore prohibited during class. There are exceptions for students with specific note-taking needs, who must notify the instructor at the beginning of the semester if they will use such devices.

Last Day to Withdraw:
In order to withdraw from a course it is not sufficient simply to stop attending class or to inform the instructor. In accordance with the policy, contact your advisor or the Registrar to begin the withdrawal process. The last day for withdrawal from this course is cited in the official catalog.

Academic Integrity Policy:
Academic dishonesty in any form will not be tolerated in this course. Cheating, copying, plagiarizing, or any other form of academic dishonesty (including doing someone else's individual assignments) will result in, at the extreme minimum, a zero on the assignment in question, and could result in a failing grade in the course or even expulsion from DigiPen.

Disability Support Services:
If students have disabilities and will need formal accommodations in order to fully participate or effectively demonstrate learning in this class, they should contact the Disability Support Services Office at (425) 629-5015 or dss[at]digipen[dot]edu. The DSS Office welcomes the opportunity to meet with students to discuss how the accommodations will be implemented. Also, if you may need assistance in the event of an evacuation, please let the instructor know.

Tentative Weekly Topics:

Week	Dates	Topics
1	Sep 4-9	Introduction to data science
2	Sep 10-16	Perceptron and PLA
3	Sep 17-21	Exploratory data analysis
4	Sep 24-28	Probability and statistics review, statistical inference
5	Oct 1-7	Spam filters & naïve Bayes
6	Oct 8-12	k-nearest neighbors
7	Oct 15-19	**Midterm exam**
8	Oct 22-26	Simple linear regression
9	Oct 29- Nov 2	Multiple regression
10	Nov 5-9	Logistic regression
11	Nov 13-16	Decision trees
12	Nov 19-21	Neural networks
13	Nov 26-30	Clustering and k-means
14	Dec 3-7	Centrality & PageRank algorithm
15	Dec 10-14	**Final exam**

This schedule is subject to change. Please check the class web page for up to date information.