Helwan University
Faculty of Computers and Information

Course Syllabus

Department of Computer Science
CS 395: "Data Science"
Fall 2017

Office Hours: See webpage.

Prerequisites: No official prerequisites to register this course. Only, minimal understanding of calculus and probability at the level of high school and elementary college are needed.

Objectives: This course looks like a "trailer" of a whole field that consists of 15+ courses. Therefore, it is neither a prerequisite to any other course nor any other course is a prerequisite to it (except ordinary calculus and probability at the level of high school or elementary college). This course should motivate students to study each course in this field and excel in it.

Text: No official text is required, since this course is a compilation of several elementary topics from different introductory courses.

Course Syllabus: The course starts with some basics of probability theory and statistics, some basics of Linear Algebra, and some basics of data visualization.

Then, the course will be covering basics of Statistical Decision Theory, from which we will merge to Linear Models and Regression. Bayes' classifier will be derived and explained, along with application to multinormal distributions. This will take us to define some statistical concepts, e.g., estimation, loss function, minimizing the risk, etc. Some basic methods, for both regression and classification, will be covered in varying levels of details, e.g., Neural Networks (NN), K-Nearest Neighbor, logistic regression, and Classification and Regression Trees (CART). Since each topic these is detailed in a stand alone course of the 15 courses, e.g., probability, statistics, pattern recognition, this course will be practical and not theoretical. We will not cover mathematical derivations nor heavy mathematical treatment. Emphasize will be on intuition and computer problems using Python.

Assignments: Assignments will include both, problems and computer exercises. Python is preferable for solving the computer exercises. **No late assignments please**.

Course Project: Every group should select one project. A group is consisted of some students. **By the 4th week** the names of the members of each group should be registered with the TA of the course. The project topic should be determined and approved before the mid term.

Grading Policy: 60% of the grade will be on the final exam, 0% on homeworks, 15% on quizes(which will be selected from homeworks), 15% on midterm, and 10% on course project. Solving assignments, in both formats the paper-and-pencil and computer exercises, is crucial for acquiring the skills to solve the exam and course project.

General Info:

- All handouts, grades, and assignments will be posted on the course webpage.
- Final exam will be in the form of Multiple Choice Questions (MCQ). Every question will have five answers, one of them is correct. Every four wrong answers cancel one correct answer. Exams will be open book. So, focus in your course on learning and understanding **NOT** on memorizing.
- For applications on real data sets, a good data repository is `http://mlearn.ics.uci.edu/MLRepository.html`