Helwan University
Faculty of Computers and Information

Course Syllabus

Department of Computer Science
Pattern Recognition

Office Hours: See webpage.

Prerequisites to Understand: No official prerequisites to register this course. However, two courses are very important prerequisites to understand this course, a course in linear algebra and a course in probability.

Objectives: This course teaches students the basics of *learning* the relationships between an output (the response variable) and the set of inputs (the predictors) in a particular problem. Two major subfields will be studied: *regression*, where the response is quantitative (has values), and *classification*, where the response is qualitative (has labels, e.g., diseased or nondiseased).

Another objective is to get the students to the level of analyzing real life data sets and designing the appropriate learning function to minimize the prediction error. This will be fulfilled through the computer exercises and the course project.

Text: The main texts of the course will be Hastie, T., Tibshirani, R., and Friedman, J. H. (2001), *The elements of statistical learning : data mining, inference, and prediction*, Springer series in statistics, New York: Springer, along with the lecture notes. However, the subject has no borders to stop at. A list of other important texts is posted on the course webpage.

Course Syllabus: The course starts with a revision of basics of probability theory; this will include some advanced topics, e.g., multivariate analysis with emphasis on multinormal distribution. Basics of Linear Algebra will be formulated, along with important topics as the four spaces of the matrix, Singular Value Decomposition (SVD) and connection to Principal Component Analysis (PCA).

Then, the course will be covering the field in a breadth-first approach. The start will be 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).

The course ends with different metrics of assessments, e.g., Receiver Operating Characteristics Curve (ROC), and Area Under the Curve (AUC). A basic methods of assessment and design is Cross Validation (CV); this will be explained and experimented.

Assignments: Assignments will include both, problems and computer exercises. Matlab 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 second 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, 10% on homeworks, and 30% on a 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