CAP 6610 – Machine Learning

Instructor: Dr. Paul Gader
Classroom: E122
Office CSE356 **Office Hours**: XXX
Phone: 392-1526 **Class Times**: MWF 4th Period, 10:40-11:30 AM
Email: pgader@cise.ufl.edu
URL: http://www.cise.ufl.edu/class/cap6610sp09/

Textbook: Pattern Recognition and Machine Learning by Christopher M. Bishop
Springer, 2006. ISBN-10:0-387-31073-8; ISBN-13: 978-0387-31073-2

Prerequisites: Mathematics for Intelligent Systems is strongly recommended. Otherwise, a good background in probability theory and linear algebra with a focus on basic concepts in probability and multivariate Gaussian, or Normal, distributions. Knowledge of MATLAB very helpful.

Topics:
The course will cover concepts involved in developing computer programs that learn, with an emphasis on methods based on probability and statistics and optimization. These techniques are widely used in language modeling, learning, and recognition applications such as speech, handwriting, and gesture recognition, computer vision, data mining, and bioinformatics. The course begins with a brief look at some fundamental problems involved in machine learning followed by a brief discussion of decision theory. Following that, an in-depth investigation of linear models for regression and classification will be undertaken. Kernel methods will be studied in detail with emphasis on Support and Relevance Vector Machines but with some discussion of other methods involving sparsity promoting priors. Some fundamental questions involving the capabilities of these systems will then be addressed in the context of Statistical Learning Theory. Mixture models, including those involved in combining models, decision trees, and the Expectation-Maximization algorithm will also be covered. Methods for performance evaluation and training, including re-substitution, cross-validation, bagging, boosting, and bolstered error estimation will be discussed.

Evaluation and Grading:
There will be 2 tests, 2 computer projects, and a cumulative final. Each test will cover a subset of the material. Homework problems, mainly from the text, will be assigned. Some of the homework will require writing simple MATLAB scripts. Your knowledge of the homework will also be evaluated by putting problems identical or very similar to homework problems on the tests. MATLAB will be the preferred implementation environment for computer projects. Students should be prepared to demonstrate their projects by running their programs if asked. Each project will involve implementing a machine learning program, running it on some data, writing a report, and possibly demonstrating your program. Grades will be determined as follows:

Projects, 20% each x 2 projects	= 40%
Tests, 15% each x 2 tests	= 30%
Final	= 20%
Homework	= 10%