CSCI 538 Artificial Intelligence
(Machine Learning and Data Analysis)

INSTRUCTOR:

Derek Harter, Ph.D., Associate Professor, Department of Computer Science
Texas A&M University – Commerce
Office: Jour 208;
E-mail: Derek.Harter@tamuc.edu
Office Hours: T, W, Th 1:30pm-5:00pm or by appointment.

CLASS MEETINGS:

01W	81764	Web Based Class

DESCRIPTION:

This semester's AI course will focus on the sub field of *machine learning* with some work in the analysis of big data. We will look at classic unsupervised and supervised learning methods, used in the field to classify data, cluster it and find optimizations. This will include looking at k-means and hierarchical clustering, self-organizing maps, linear regression, decision trees, optimization techniques such as genetic algorithms, etc. We will cover ways of getting hold of interesting datasets, ideas on how to collect data from users, and many different ways to analyze and understand the data once you've found it.

REQUIREMENTS AND OBJECTIVES:

- Explore downloading and mining real web data sets.
- Learn about unsupervised methods for grouping and visualization.
- Program optimization algorithms to search for optimal solutions using hill climbing, GA's, etc.
- Become familiar with some advanced classification techniques of the kernel methods family of algorithms.

COMPANION TEXTBOOKS:

[CI] *Programming Collective Intelligence* by Toby Segaran, O'Reilly Media, 2007, ISBN: 0596529325

[ML] Marsland. (2009). Machine Learning: An Algorithmic Perspective. By Marsland, CRC Press, 2009.

[PY] *How to Think Like a Computer Scientist: Learning with Python 2ed* by Jeffrey Elkner, Allen B. Downey and Chris Meyers (Open Book Project)
http://www.greenteapress.com/thinkpython/

PREREQUISITES:

Background in basic programming (Undergraduate equivalent of CSci 151/152) and Data Structures (Undergraduate equivalent of CSci 270) or equivalent knowledge will be assumed for this course.

EVALUATION:

Your grade for the course will be based on the following (approximate) percentages:

Two Exams (Midterm and Final):	50%
Programming Assignments (approximately 4):	45%
Class Participation and Attendance:	5%

STUDENTS WITH DISABILITIES ACT COMPLIANCE:

The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires that all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation of their disabilities. If you have a disability requiring an accommodation, please contact: Office of Student Disability Resources and Services, Texas A&M University-Commerce, Gee Library, Room 132, Phone (903) 886-5150, StudentDisabilityServices@tamuc.ed

ACADEMIC ETHICS:

"All students enrolled at the University shall follow the tenets of common decency and acceptable behavior conducive to a positive learning environment." (See Student's Guide Handbook, Policies and Procedures, Conduct).

ATTENDANCE POLICY:

Students are expected to follow all instructions and visit eCollege regularly many times weekly to complete the materials for this online course. If a student is unable to submit assignments by the due date for the assignment, they are expected to make alternative arrangements to assure that the assignment is turned in ON TIME, before the assignment is actually due. Any student wishing to withdraw from the course must do so officially as outlined in the class schedule. THE INSTRUCTOR CANNOT DROP OR WITHDRAW ANY STUDENT.

COURSE REQUIREMENT DEADLINES:

Credit will be given for ONLY those exams, programs, and/or projects turned in no later than the deadline as announced by the instructor of this class, unless prior arrangement has been made with the instructor.

TENTATIVE SCHEDULE

Wk	Date	Topic / Activity	Notebook	Assg
1	08/25	Course Introduction	1a & 1b Intro to Python Programming	
		Set up Python		
2	09/01	Introduction to Python Scientific Libraries	2a & 2b Numpy and Matplotlib	#1
3	09/08	Principle Component Analysis (Uns)	3 - PCA	
4	09/15	K-means and Hierarchical Clustering (Uns)	4 - KMeans	
5	09/22	Linear Regression & Perceptrons (Sup)	5 - Linear Regression	
				#2
6	9/29	Classification (Sup)	6 - Classification	
7	10/06	Logistic Regression (Sup)	7 - Logistic Regression	
8	10/13	**Exam 1**		
9	10/20	K Nearest Neighbor (Sup)	8- K Nearest Neighbors	#3
10	10/27	Decision Trees and Ensembles (Sup)		
11	11/03	Optimization & Search (Sup)	10 - Optimization	
12	11/10	Genetic Algorithms (Sup)		
				#4
13	11/17	Support Vector Machines (SVM) and Kernel Methods (Sup)		
14	11/24			
		Thanksgiving		
15	12/01	Naive Bayes (Sup)		
Fnl	12/08-12/13	**Exam 2**		

Student Learning Outcomes

1. Develop familarity with high-level Python scripting language
2. Learn basics of fundamental machine learning techniques, such as optimization, bayesian estimates, clustering, k-nearest neighbor, kernel methods, etc.
3. Learn basic distinction between supervised and unsupervised machine learning methods.
4. Show examples of using Web 2.0 data sources for systems development.
5. Learn basic machine learning training and testing techniques, including cross validation and data optimization.

Learning outcomes will be measured through mapping assignment and test questions to specific outcome items, as well as through exit surveys of student experiences with the outcome familarity.