CPSC 689-607: Special Topics in Pattern Classification
Fall 2008
Time: TR 11:10AM-12:25PM, **Room**: SCTS 216

Instructor:	Ricardo Gutierrez-Osuna
Office:	520A HRBB
Office Hours:	TR 12:30-1:00 PM
Phone:	979.845.2942
Email:	rgutier@cs.tamu.edu
URL:	http://faculty.cs.tamu.edu/rgutier/

Catalog Description: Selected topics in an identified area of computer science.

Detailed Description: Introduction to computational models for the analysis, classification and clustering of patterns in high-dimensional spaces. The course covers fundamental concepts, including decision theory, density estimation, dimensionality reduction, statistical classification and clustering, as well as advanced topics such as manifold learning, statistical learning theory, hidden Markov models, and ensemble learning.

Prerequisites: CPSC 206, MATH 222, MATH 411 (or equivalent) and graduate standing in CPSC, CECN, ELEN, CEEN (or permission of the instructor). Basic knowledge of Linear Algebra, Probability and Statistics: algebra of matrices, geometry of Euclidean space, vector spaces and subspaces, basis, linear independence, linear transformations, eigenvalues and eigenvectors, mean, variance, probability and distributions. Programming experience in a high-level language is required.

Textbook:
- C. M. Bishop, Pattern Recognition and Machine Learning, 1st ed., Springer, 2006.

Recommended:
- D. Hanselman and B. Littlefield, Mastering MATLAB 7, Prentice Hall, 2004.

References:
- A. R. Webb, Statistical Pattern Recognition, 2nd ed., Wiley, 2002.
- R. O. Duda, P. E. Hart and D. G. Stork, Pattern Classification, 2nd ed., Wiley, 2001.
- K. Fukunaga, Introduction to Statistical Pattern Recognition, 2nd ed., Academic Pr., 1990.

Course Objectives: The objectives of this course are to:
- Introduce the fundamental concepts of pattern recognition
- Provide the students with a toolbox of methods and algorithms they can use for practical pattern recognition problems

Course Outcomes: Upon satisfactory completion of the course, the student will be able to:
- Analyze a pattern recognition problem and present a valid formulation
- Propose and evaluate possible methods to solve the problem
- Implement a number of algorithms in a high-level language
- Design an experiment to validate formulation and implementation

Course Outline

- Introduction (3 lectures)
  - Pattern Recognition
  - Probability and Statistics
  - MATLAB® and Linear Algebra
  - Fourier analysis
- Pattern Classifiers (7 lectures)
  - Bayesian Decision Theory
  - Quadratic Classifiers
  - Parameter and Density Estimation
  - Nearest Neighbors
  - Linear Discriminant Functions
  - Cross-Validation
- Feature and Dimensionality Reduction (4 lectures)
  - Principal Components Analysis
  - Fisher's Discriminants Analysis
  - Feature Subset Selection
  - Advanced methods (snapshot PCA, oriented PCA, NMF, LLE, ISOMAP)
- Clustering and Unsupervised Learning (4 lectures)
  - Mixture models and Expectation Maximization
  - Statistical Clustering
  - Independent Components Analysis
- Advanced Topics (7 lectures)
  - Support Vector Machines
  - Kernel PCA/LDA
  - Hidden Markov Models
  - Ensemble Learning

Grading: The course grade will be the weighted sum of four grades. Grading will be straight scale (90-100 A, 80-89 B, 70-79 C, 60-69 D, below 60 F). These numeric thresholds may be lowered due to clustering, but will not be raised.

- **Homework**: There will be three homework assignments, distributed every 2-3 weeks during the first part of the semester. Homework assignments will emphasize the implementation (programming) of material presented in class. *Homework assignments must be done individually*.
- **Tests**: There will be a midterm exam and a final exam. All tests will be closed-books, closed--notes. One double-sided, hand-written sheet (8.5 x 11") will be allowed. Tests will have an emphasis on new material from the class notes or the reading assignments.
- **Project**: The last part of the semester will be dedicated to a term project. Students are encouraged to propose projects related to their own research. The projects must be performed in groups of up to three people. Projects will be graded by their content (75%) and the quality of a classroom presentation (25%) at the end of the semester. Peer reviews will also be performed to measure the individual members' participation in the project. G*rading criteria for the project presentation, final report and peer reviews are available in the course webpage.*

	Weight (%)
Homework	40
Project	30
Midterm	15
Final Exam	15

Homework submissions. Homework assignments are due at *11:00AM* on the due date. Electronic material will be submitted with the "*turnin*" utility at https://csnet.cs.tamu.edu; hardcopies will be submitted directly to the instructor. Email submissions will not be accepted. Note that '*turnin*' has a maximum file size that can be submitted. With the exception of MATLAB's built-in functions (e.g., cov, eig, mvnrnd), you are expected to write your own implementation of the algorithms; in case of doubt please consult with the instructor.

Late submissions. Late submissions (i.e., as flagged by *csnet*) will receive a 15% penalty on the total grade of the assignment; the penalty will increase by an additional 15% every 24 hours. Hardcopies of late submissions must be *date and time stamped* by the staff in the Computer Science main office. An assignment is considered submitted when ALL components of the assignment have been submitted; e.g., late submission of one problem in a homework will cause your entire homework to be considered as a late submission.

Missed Tests: Missed tests can only be made up in case of emergency or work conflicts, and will require supporting documentation. Whenever possible, these issues should be discussed with the instructor prior to the conflicting date.

Collaboration vs. Academic Dishonesty: Students are encouraged to exchange ideas and form study groups to discuss the course material, and prepare for homework assignments and tests. However, discussions regarding homework assignments should be kept at the conceptual level (i.e., sharing code is not allowed). Scholastic dishonesty will not be tolerated in homework assignments, tests or projects. For a list of examples of scholastic dishonesty see Section 20 of the TAMU Student Rules (http://student-rules.tamu.edu/).

Academic Integrity Statement

"*An Aggie does not lie, cheat, or steal or tolerate those who do.*" Please review the Aggie Honor Code and Honor Council Rules and Procedures at http://www.tamu.edu/aggiehonor.

Course Schedule

Week / day		Topics	Supplemental reading	Assignments / activities
1	8/26	Introduction to Pattern Recognition	Section 1.1	
	8/28	Probability and Linear Algebra	1.2, 2.3, Appendix C	HW #1 assigned
2	9/2	Fourier Analysis		
	9/4	Bayesian Decision Theory	1.5	
3	9/9	Quadratic Classifiers	4.2	
	9/11	Parameter Estimation		
4	9/16	Kernel Density Estimation	2.5.1, 3.2	
	9/18	Nearest Neighbors	2.5.2	HW #1 due; HW #2 assigned
5	9/23	Linear Discriminant Functions	4.1.1 - 4.1.3, 4.1.7	
	9/25	Cross-validation	1.3	
6	9/30	Principal Components	1.4, 12.1, Appendix E	
	10/2	Fisher Linear Discriminants	4.1.4 - 4.1.6	
7	10/7	Feature Subset Selection		Homework #2 due
	10/9			**Midterm Exam**
8	10/14	Advanced dimensionality reduction	Roweis; Tenenbaum; 12.1.4, 12.4.3	Homework #3 assigned
	10/16	Mixture Models and EM	9.2-9.4	
9	10/21	Statistical Clustering	9.1	
	10/23	Independent Components Analysis	Hyvarinen; 1.6, 12.4.1	
10	10/28	Support Vector Machines	7, Burges	Homework #3 due
	10/30	SMVs and Kernel Methods	7, Burges	Term project proposal due
11	11/4	Kernel Methods	7, Burges	
	11/6	Kernel PCA, LDA and regression	12.3, Mika	
12	11/11	Discrete HMMs, Viterbi	13.1-2, Rabiner	
	11/13	Baum-Welch, Continuous HMMs	13.1-2, Rabiner	
13	11/18	Ensemble Learning	14	
	11/20			Catch-up / review day
14	11/25			Project presentations I
	11/27	Thanksgivings (no class)		
15	12/2			Project presentations II
	12/4			Reading day (no class)
	12/5			**Final Exam 3:00-5:00PM**