CSCE 666: Pattern analysis
Spring 2020

Time: MWF 3:00-3:50pm, **Room**: HRBB 126

Instructor: Ricardo Gutierrez-Osuna
Office: 506A HRBB
Office Hours: Fridays 3:50-4:20pm or by appointment
Phone: 979.845.2942
Email: rgutier@tamu.edu

Catalog description: Introduction to methods for the analysis, classification and clustering of high-dimensional data in Computer Science applications. Course contents include density and parameter estimation, linear feature extraction, feature subset selection, clustering, Bayesian and geometric classifiers, non-linear dimensionality reduction methods from statistical learning theory and spectral graph 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 (4 lectures)
  - Pattern pecognition
  - 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 (3 lectures)
  - Mixture models and Expectation Maximization
  - Statistical clustering
  - Independent components analysis
- Advanced Topics (6 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' contributions to 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 *8:00AM* 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 Code of Honor at http://student-rules.tamu.edu/aggiecode

Course schedule

Week/day		Special events	Topics (supplemental reading)
1	1/13		Introduction to pattern recognition (Section 1.1) Probability and linear algebra (1.2, 2.3, Appendix C)
	1/15	HW1 assigned	
	1/17		
2	1/20		Fourier analysis Bayesian decision theory (1.5)
	1/22		
	1/24		
3	1/27		Quadratic classifiers (4.2) Parameter estimation
	1/29		
	1/31		
4	2/3		Kernel density estimation (2.5.1, 3.2) Nearest neighbors (2.5.2)
	2/5	HW1 due; HW2 assigned	
	2/7		
5	2/10		Linear discriminant functions (4.1.1 - 4.1.3, 4.1.7) Cross-validation (1.3)
	2/12		
	2/14		
6	2/17		Principal components (1.4, 12.1, Appendix E) Fisher linear discriminants (4.1.4 - 4.1.6)
	2/19		
	2/21		
7	2/24	HW2 due	Feature subset selection
	2/26	Midterm Exam	
	2/28		
8	3/2	HW3 assigned	Advanced dimensionality reduction (Roweis; Tenenbaum; 12.1.4, 12.4.3) Mixture models and EM (9.2-9.4)
	3/4		
	3/6		
9	3/9	Spring break	
	3/11	Spring Break	
	3/13	Spring Break	
10	3/16		Statistical clustering (9.1) Independent components analysis (Hyvarinen; 1.6, 12.4.1)
	3/18		
	3/20		
11	3/23	HW3 due	Support vector machines (7, Burges)
	3/25	Proposal presentations	
	3/27		
12	3/30		SMVs and kernel methods (7, Burges) Kernel methods (7, Burges)
	4/1		
	4/3		
13	4/6		Kernel PCA, LDA and regression (12.3, Mika) Discrete HMMs, Viterbi (13.1-2, Rabiner)
	4/8		
	4/10	Reading day (no class)	
14	4/13		Baum-Welch, continuous HMMs (13.1-2, Rabiner) Ensemble learning (14)
	4/15		
	4/17		
15	4/20		Ensemble learning
	4/22		
	4/24		
16	4/27	Final exam	
	4/29	Reading day (no class)	
	5/1	No class	
17	5/4	Project presentations 10:30am – 12:30pm	