STA4634/5635 Applied Machine Learning
Fall 2018

Course Information
Class Meeting Place: HCB 215
Class Meeting Time: MW 5:15-6:30pm

Instructor: Dr. Adrian Barbu
E-mail: abarbu@stat.fsu.edu
Office: 106C OSB
Phone: 850-290-5202
Office Hours: Tuesday 3:00-5:00pm or by appointment

Teaching Assistant: Zishen Xu
E-mail: zx16@my.fsu.edu
Office: 204 OSB
Office Hours: Monday/Wednesday 4:00-5:00pm

Textbooks (optional):
1. **The Elements of Statistical Learning** by T. Hastie, R. Tibshirani, and J. H. Friedman (publisher: Springer) http://www.stanford.edu/~hastie/ElemStatLearn/printings/ESLII_print12.pdf
2. **Pattern Recognition and Machine Learning** by Christopher M. Bishop (publisher: Springer)
3. **Machine Learning** by Tom M. Mitchell (publisher: McGraw-Hill)

All textbooks are **optional** since the course will not follow any particular book.

Course Objectives: At the end of the course, the student will:
- be able to understand many machine learning methods with their advantages and disadvantages
- be able to implement the methods or know where to obtain them from
- be able to use existing library software
- have a working knowledge of most of the methods
- be able to determine most appropriate learning method for a specific application

Course topics: This course is an overview of statistical methods for supervised, unsupervised and weakly supervised learning. The following topics will be covered:
- Decision Trees, Random Forests
- Naive Bayes Classifiers
- Linear and Logistic Regression
- Generative and Discriminative Learning
- Learning with regularized loss functions
- Neural Networks
- Large Margin Classifiers: Support Vector Machines, Kernel Methods
- Boosting: AdaBoost, LogitBoost, RealBoost, GentleBoost
- Feature Selection with Annealing

- Efficient Inference: Marginal Space Learning
- Learning Issues: Overfitting, Bias-variance tradeoff
- Learning Theory: PAC learning, VC Dimension
- Graphical Models, Hidden Markov Models, Conditional Random Fields, Belief Propagation
- Semi-supervised Learning
- Unsupervised Dimensionality Reduction: PCA, Factor Analysis, ICA
- Supervised dimensionality reduction: Feature Selection, Fisher LDA, Hidden layers in NN
- Nonlinear Dimensionality Reduction: Kernel PCA, Multi-dimensional scaling (MDS), Isometric mapping (ISOMAP), Local linear embedding (LLE)
- Maximum Entropy models: FRAME
- Using Incomplete Data: MLE and EM
- Unsupervised learning: K-means, EM, Spectral clustering, Self Organizing Maps
- Reinforcement Learning
- Metric Learning

For each method, examples from different fields such as Natural Language Processing, Bioinformatics, Computer Vision, and Medical Imaging will be presented. Some of the most important methods will accompanied by small projects for a better understanding of their advantages and limitations.

Projects (capped at maximum 90 points total):

	Project	Needs Programming	Points	Due
1	Decision Trees	Not really	8	09/05
2	Random Forest	Not really	8	09/12
3	Logistic Regression	Yes	8	09/19
4	TISP	Yes	8	09/26
5	Weka	No	10	10/03
6	FSA regression and binary clf	Yes	10	10/17
7	FSA multi-class	Yes	15	10/24
8	Boosting	Yes	10	10/31
9	Clustering	Yes	8	11/07
10	Neural Nets/CNN	Yes	10	11/14
11	HMM	Yes	10	11/28
12	PCA	Yes	8	12/05

Grading: There will be 12 homework projects shown above worth at most 90 points, and attendance worth another 10 points for a total of 100 points.

- The projects are worth at most 90 points. Students can choose which projects to work on to reach 90 points. If they obtain more than 90 points for the projects, only 90 points will be counted towards the final grade.
- The following scheme will be used to convert the percentage points to letter grades

[90, 95)	A-	[95, 100]	A		
[80, 83)	B-	[83, 87)	B	[87, 90)	B+
[70, 73)	C-	[73, 77)	C	[77, 80)	C+
[60, 63)	D-	[63, 67)	D	[67, 70)	D+
[0, 60) F					

Information on the datasets and their training and testing sets

Dataset	Type	Obs	Features	Train	Test
Arcene	Binary clf	100+100	10000	train	valid
Dexter	Binary clf	300+300	20000	train	valid
Dorothea	Binary clf	800+350	100000	train	valid
Gisette	Binary clf	6000+1000	5000	train	valid
Hill-valley	Binary clf	606+606	100	X,Y	Xtest,Ytest
Madelon	Binary clf	2000	500	train	valid
Miniboone	Binary clf	130k	50	80/20 random splits	
Wilt	Binary clf	4,889	6	train	test
Covtype	Multi-class clf	580k	54	first 11,340 + next 3,780	last 565,892
Poker	Multi-class clf	25k+1mil	10	X,Y	Xtest,Ytest
Satimage	Multi-class clf	4435+2000	36	X,Y	Xtest,Ytest
Abalone	Regression	4177	8	80/20 random splits	
Bike rental	Regression	11k+6.5k	10	train	test+online
Online News	Regression	40k	58	80/20 random splits	

Prerequisites: STA 3032 and knowledge of Matlab, R, Python, C++ or other programming language or consent of instructor.

Course Materials

- CMU Machine Learning Class: http://www.cs.cmu.edu/~epxing/Class/10701/
- Trevor Hastie's ML books: http://www.stanford.edu/~hastie/pub.htm
- Tom Michell's ML book website: http://www.cs.cmu.edu/~tom/mlbook.html
- Nillson's ML book: http://ai.stanford.edu/~nilsson/mlbook.html
- Blackboard class website: go to http://campus.fsu.edu/ and login using you ACNS username and password. Homework, datasets, grades, course notes and other course material will be posted there.

Course Policy

- **Classroom policies:** The classroom environment is an important factor for effective learning. In order to not distract other students' attention please follow these classroom policies. The first one of these is the university policy.
    - Remember that no food or drinks are allowed in the classroom.
    - Turn off all audible alarms (cell phones, pagers, calculators, watches etc.)
    - Do not use cell phones in the class.
    - Come to the class on time. Opening and closing the classroom door in the middle of a class cause distraction to the students and the teacher.
    - Do not talk to other students without permission while the professor is teaching. More than one conversation creates noise and makes it difficult for the students to pay attention to the lecture.
- **Homework:** There will be 12 homework projects, due one to two weeks from the date they are announced. The homework must be neatly written, preferably typed and **must be submitted online**. Computer output should be kept to a minimum. You are encouraged to submit the project code by email. The code for best results for each homework will be posted on Blackboard to be available for all students attending the class. Students are allowed to work on the projects in teams of two (for

graduate students) and three (for undergrads) and should **submit a single homework for each team**.

- **Code:** It is acceptable to use code downloaded from the internet for the homework as long as a reference to the code website, package or the appropriate paper is added to the homework report.
- **Collecting returned homework:** It is the student's responsibility to check grades on the Blackboard class page. If you notice any mistake in recording grades on the Blackboard page, please inform the instructor about it as soon as possible.
- **Homework re-grade:** You have one week to request a re-grade of a homework from the date on which the graded homework is returned to the students of the class. For that, see the instructor along with the relevant homework.
- **Contacting the instructor outside the class:** You are strongly encouraged to come to the instructor during his office hours. If your schedule conflicts with the office hours, you can make an appointment. You may ask the instructor brief questions by e-mail, but you may be asked to come to office hours if the instructor thinks that the questions are better answered in person.
  When you send e-mails remember the following:
  - Always e-mail from your FSU accounts. The e-mails from non-FSU accounts may not reach me due to filters.
  - Always write your full name at the end of each e-mail message you send.
  - Always write the course number STA 5635 at the beginning of the subject line.
- **University Attendance Policy:** Excused absences include documented illness, deaths in the family and other documented crises, call to active military duty or jury duty, religious holy days, and official University activities. These absences will be accommodated in a way that does not arbitrarily penalize students who have a valid excuse. Consideration will also be given to students whose dependent children experience serious illness.
- **Academic honor policy:** The Florida State University Academic Honor Policy outlines the University's expectations for the integrity of students' academic work, the procedures for resolving alleged violations of those expectations, and the rights and responsibilities of students and faculty members throughout the process. Students are responsible for reading the Academic Honor Policy and for living up to their pledge to ". . . be honest and truthful and . . . [to] strive for personal and institutional integrity at Florida State University." (Florida State University Academic Honor Policy, found at http://dof.fsu.edu/honorpolicy.htm.)
- **Americans with Disabilities Act**:
  Students with disabilities needing academic accommodation should:
  1) register with and provide documentation to the Student Disability Resource center; and
  2) bring a letter to the instructor indicating the need for accommodation and what type.
  This should be done during the first week of class.
  This syllabus and other class materials are available in alternative format upon request.

For more information about services available to FSU students with disabilities, contact:

Student Disability Resource Center
874 Traditions Way
108 Student Services Building

Florida State University
Tallahassee, FL 32306-4167
(850) 644-9566 (voice)
(850) 644-8504 (TDD)
sdrc@admin.fsu.edu
http://www.disabilitycenter.fsu.edu/

- **Syllabus Change Policy**
  Except for changes that substantially affect implementation of the evaluation (grading) statement, this syllabus is a guide for the course and is subject to change with advance notice.