STA5635 Applied Machine Learning
Fall 2022

Course Information

Class Meeting Place: **Online**
Class Meeting Time: Tu/Th 4:50-6:05pm

Students all meet with instructor for class weekly, synchronously, using Zoom, at a specific class time indicated in the University's course schedule for each semester. Additional asynchronous interactions (e.g., discussion forums) among students and with instructor may also be required to complete the course.

Instructor: Dr. Adrian Barbu

E-mail: abarbu@stat.fsu.edu
Office: 305 OSB
Phone: 850-290-5202
Office Hours: Wednesday 3:30-5:30pm (on Zoom) or by appointment

Teaching Assistant: Cheng Long

E-mail: : cl18t@fsu.edu
Phone: 908-873-9957
Office: 310 OSB
Office Hours: Wednesdays 1:00-3:00pm
Zoom link for Office Hours: https://fsu.zoom.us/j/95632235948

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.

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

Course Objectives: At the end of the course, the student will be able to:

- compare and contrast many machine learning methods with their advantages and disadvantages
- implement the methods or know where to find their implementation
- use existing library software
- determine the 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
- Learning Issues: Overfitting, Bias-variance tradeoff
- Learning Theory: PAC learning, VC Dimension
- Graphical Models, Hidden Markov Models, Conditional Random Fields, Belief Propagation
- Unsupervised Dimensionality Reduction: PCA, Factor Analysis, ICA
- Supervised dimensionality reduction: Feature Selection, Fisher LDA
- Nonlinear Dimensionality Reduction: Kernel PCA, Multi-dimensional scaling (MDS), Isometric mapping (ISOMAP), Local linear embedding (LLE)
- Using Incomplete Data: MLE and EM
- Unsupervised learning: K-means, EM, Spectral clustering, Self Organizing Maps
- Semi-supervised Learning
- 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:

	Project	Needs Programming	Points
1	Decision Tree/Random Forest	Not really	10
2	Regression	Yes	7
3	Logistic Regression	Yes	8
4	TISP	Yes	7
5	Weka	No	10
6	FSA regression and binary clf	Yes	9
7	Boosting	Yes	9
8	Neural Nets/CNN	Yes	10
9	HMM	Yes	9
10	Clustering	Yes	7
11	PCA	Yes	7
12	TBA	Yes	10

Grading: There will be 12 homework projects shown above worth at most 89 points, and quizzes worth another 14 points for a total of 103 points. There will be extra 3 bonus points given at the discretion of the instructor for students that have actively participated in the class discussions.

- The largest 10 project grades will be considered, for a total of at most 89 points.
- The following scheme will be used to convert the percentage points to letter grades

[90, 93)	A-	[93, 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+600	500	train	valid
Miniboone	Binary clf	130k	50	80/20 random splits	
Wilt	Binary clf	4339+500	5	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	

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
- Canvas class website: go to http://canvas.fsu.edu/ and login using your FSUID and password. Homework, datasets, grades, course notes and other course material will be posted there.

Course Policy

- **Individual study:** You are expected to read the course material beforehand and ask questions that you have in class.
- **Discussion sessions:** Every class will contain a discussion session where students are expected to participate, present what they understood and ask any questions that they might have about the class material or the homework.
- **Quizzes:** There will be weekly online quizzes on Canvas to check whether the students have studied the material before class. **You must work on the quizzes yourself and not have anybody else do them for you**.
- **Homework:** There will be 12 homework projects, due one to two weeks from the date they are announced. The homework must be written in a reproducible notebook format such as Rmarkdown or Jupyter for Python, and **must be submitted online**. Computer output should be kept to a minimum. Students are allowed to work on the projects in teams of two 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 quizzes/homework:** It is the student's responsibility to check grades on the Canvas class page. If you notice any mistake in recording grades on the Canvas 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 for an online or in person meeting. You may ask the instructor brief questions by e-mail, but you may be asked to come to office hours or meet in a video-conference session 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://fda.fsu.edu/Academics/Academic-Honor-Policy)

- **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.
  Please note that instructors are not allowed to provide classroom accommodation to a student until appropriate verification from the Student Disability Resource Center has been provided.

  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 the:

  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/

- **Free Tutoring from FSU:** On-campus tutoring and writing assistance is available for many courses at Florida State University. For more information, visit the Academic Center for Excellence (ACE) Tutoring Services' comprehensive list of on-campus tutoring options at http://ace.fsu.edu/tutoring or contact tutor@fsu.edu. High-quality tutoring is available by appointment and on a walk-in basis. These services are offered by tutors trained to encourage the highest level of individual academic success while upholding personal academic integrity.

- **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.