**CS7646 ML4T**
**Machine Learning**
**for Trading**

# MACHINE LEARNING FOR TRADING COURSE

## FALL 2020 SYLLABUS

## OVERVIEW

This course introduces students to the real world challenges of implementing machine learning based trading strategies including the algorithmic steps from information gathering to market orders. The focus is on how to apply probabilistic machine learning approaches to trading decisions. We consider statistical approaches like linear regression, Q-Learning, KNN and regression trees and how to apply them to actual stock trading situations.

This course is composed of three mini-courses:

- Mini-course 1: Manipulating Financial Data in Python
- Mini-course 2: Computational Investing
- Mini-course 3: Machine Learning Algorithms for Trading

A set of course notes and example code can be found here: [[1]]

## VIDEO CONTENT

The official video content for this course is available on Canvas and also for free at Udacity.

## IMPORTANT NOTE

This course ramps up in difficulty towards the end. The projects in the final 1/3 of the course are challenging. Be prepared.

# INSTRUCTOR INFORMATION

Tucker Balch, Ph.D.

Professor, Interactive Computing at Georgia Tech

CS 7646 Course Designer

CS 7646 Instructor: Spring 2016, Fall 2016, Spring 2017, Summer 2017 (online), Fall 2017, Spring 2018, Summer 2018, Fall 2018

CIOS reviews Media:2017SummerCIOS.pdf

Maria Hybinette

Associate Professor, Computer Science, University of Georgia

CS 4646 Instructor: Summer 2018

David Byrd

Research Scientist, Interactive Media Technology Center at Georgia Tech

CS 7646 On Campus Instructor: Summer 2016, Summer 2017, Spring 2018, Fall 2019

CS 7646 Head TA: Spring 2016, Fall 2016, Fall 2017

David Joyner

CS 7646 Online Instructor: Spring 2019, Summer 2019, Fall 2019, Spring 2020, Summer 2020

Joshua Fox

CS 7646 Co-Instructor: Fall 2020

CS 7646 Head TA: Fall 2019, Spring 2020, Summer 2020

# TEXTBOOKS, SOFTWARE & OTHER RESOURCES

We will use the following textbooks:

- For Mini-course 1: Python for Finance by Yves Hilpisch amazon.com (optional)
- For Mini-course 2: What Hedge Funds Really Do by Romero and Balch amazon.com (required)
- For Mini-course 3: Machine Learning by Tom Mitchell (optional)
  - Buy it for $218.00 at: amazon.com
  - Buy a paperback version for $61.78. IMPORTANT WARNINGS: 1) They only ship to the US 2) It takes them 3 weeks to print the book. If you order from outside the US they will quietly accept your money but never ship the book: less expensive version at mcgraw hill
  - Buy a paperback international version for $19.10. I am not certain about the reliability of this company: international

Software:

- Follow these instructions to set up the software: ML4T_Software_Setup

Other resources:
- Course notes developed by Octavian Blaga [docs.google.com]
- Pandas documentation: [pandas.pydata.org]
- David Byrd's slides on how to vectorize technical analysis methods:
  media:CDB_vectorize_me.pptx

# PREREQUISITES/CO-REQUISITES

All types of students are welcome! The Machine Learning topics might be "review" for CS students, while finance parts will be review for finance students. However, even if you have experience in these topics, you will find that we consider them in a different way than you might have seen before, in particular with an eye towards implementation for trading.

If you answer "no" to the following questions, it may be beneficial to refresh your knowledge of the prerequisite material prior to taking CS 7646:

- Do you have a working knowledge of basic statistics, including probability distributions (such as normal and uniform), calculation and differences between mean, median and mode
- Do you understand the difference between geometric mean and arithmetic mean?
- Do you have strong programming skills? Take this quiz compinvesti-prog-quiz if you would like help determining the strength of your programming skills.
- Are you competent with the Unix command line?

Who this course is for: The course is intended for people with strong software programming experience and introductory level knowledge of investment practice. A primary prerequisite is an interest and excitement about the stock market.

Software we'll use: In order to complete the programming assignments you will need to a development environment that you're comfortable with. We use Unix, but you can also work with Windows and Mac OS environments. You must download and install a set of Python modules to your computer (including NumPy, SciPy, and Pandas).

How to install the software: ML4T Software Setup

## LOGISTICS

- All course content for the course is delivered via Canvas. To access it, go to Canvas, click this course, and then click Start Here to get started!
- We will use Canvas for ALL report submissions: To access it, go to Canvas, click this course, click assignments.
- We will use Gradescope for ALL code submissions: To access it, go to Canvas, click this course, click Gradescope.
- We will use Piazza for interaction and discussion. To access it, go to Canvas, click this course, click Piazza.

## GRADING

- A: 90.0% and above
- B: 80.0% and above
- C: 70.0% and above
- D: 60.0% and above
- F: below 60.0%

Students taking the course Pass/Fail must earn at least a 75% to pass.

We do not encourage "audit" students. If you are in the course on audit status, you must earn at least a "B" on the midterm.

See semester syllabus for assignment weights.

## MINIMUM TECHNICAL REQUIREMENTS

- Browser and connection speed: An up-to-date version of Chrome or Firefox is strongly recommended. We also support Internet Explorer 9 and the desktop versions of Internet Explorer 10 and above (not the metro versions). 2+ Mbps recommended; at minimum 0.768 Mbps download speed.

- Hardware: A computer with at least 4GB of RAM and CPU speed of at least 2.5GHz.

- For code development and testing, these three configurations will work
  - PC: Windows XP or higher with latest updates installed
  - Mac: OS X 10.6 or higher with latest updates installed

- Linux: Any recent distribution that has the supported browsers installed

- For online test taking (proctortrack) you will need one of:
  - PC: Windows XP or higher with latest updates installed
  - Mac: OS X 10.6 or higher with latest updates installed
  - Linux is NOT supported.

# OFFICE HOURS

To be determined.

# PLAGIARISM

In most cases I expect that all submitted code will be written by you. I will present some libraries in class that you are allowed to use (such as pandas and numpy). Otherwise, all source code, images and write-ups you provide should have been created by you alone.

If we discover that you have submitted assignment material created by another student, either from a previous semester or in the current session, you will be assigned a 0 for the relevant project.

# CLASS POLICIES

- For Pass/Fail students: Your overall grade must be 75% or higher to get a passing grade.
- Official communication is by email: We use piazza for discussions, but it is not an official communications channel.
- Student responsibilities: Be aware of the deadlines posted on the schedule. Start work on projects even if they are not open on Canvas.
- Grade contest period: After a project grade is released you have 7 days to contest the grade. After that time projects will not be reevaluated. You must have a very specific issue with a compelling argument as to why your grade is incorrect. Example compelling argument: "The TA took 10 points off because I was missing a chart, but the chart is visible on page 5." Example not compelling argument: "I think I should have gotten more points, please regrade my project."
- Grade contest process: Instruction to be released prior to Project 1 grade release.

- Late policy: Late assignments will not be graded unless a *prior* arrangement has been made with the instructors. See CS7646_Fall_2020#Late_Work
- Exam scheduling: Exams will be held on specific days at specific times. If there is an emergency or other issue that requires changing the date of an exam for you, you will need to have it approved by the Dean of Students. You can apply for that here: http://www.deanofstudents.gatech.edu (under Resources -> Class Absences)
- Each project for this course has it's own page on this wiki. That description includes a list of specific deliverables and usually a rubric. Be sure to double check your submission against those so you don't miss anything.
- ~~Many of the projects will be revised somewhat. They are finalized when they are released on Canvas.~~
- We require that your code run properly on one of the servers we have set up at GT.
- If a problem exists with your submitted code we will not consider reassessing it if it has not been tested as described above.
- Most projects will be accompanied with template code and grading code that you can use to test your project. It is necessary that your code passes the grading checks we provide, but the final batch tests may be more rigorous. Be sure to examine the rubrics in the project description to be sure your code meets them.
- Once you are satisfied with your code, submit the EXACT same working code via Canvas.
- It is a good idea to submit a version of your working code early (before the deadline) in case some problem arises with your internet connection or Canvas.
- The latest timestamp on any part of your submission will be used as the time of submission for your whole project. Accordingly, do not resubmit anything after the deadline, or it will be considered late.
- After the submission deadline we will test your code on one of our servers which is configured identically to the ones available for your test.