ECE6254 FALL 2023
Statistical Machine Learning

Instructor: Prof. Bolin (Bo) Gao

Contact: bgao@gtsi.edu.cn
Office hour: Wednesday and Friday after 1 pm, Room 410
Meeting times: Every Wednesday and Friday starting August 21, 5:15 pm - 6:30 pm, Room 406
We will have a class on Sunday, October 8, 5:15 - 6:30 pm, but no class on Friday, September 29
Course website: https://gatech.instructure.com/courses/377204
Discussion board: https://piazza.com/class/lluauqhphof465

Welcome to ECE6254 at Georgia Tech-Shenzhen!

The impact of machine learning (ML) are being felt in every aspect of our lives. In this course, we seek to provide students with a rigorous mathematical understanding of many of the key concepts in this exciting area. The first part of the course will be focused on supervised learning, that is, ML systems that are constructed ("learned") through labeled examples. In the second part of the course, we will discuss ways of doing ML without labeled examples. The emphasis of this course is on viewing ML as systems of interconnected components. This approach will assist us in understanding the hidden mechanism behind many complex modern technologies. The goal of this course is for you to gain a deep theoretical understanding of ML as well as hands-on experience through programming exercises and a team-based project.

Prerequisites

The student should have basic fluency in **multivariable calculus, linear algebra, and probability**. We will review these mathematical concepts at various points of the course. The student should also have had experience in one or more programming language(s). The assignments of this course will be based on Python, which is the most widely used language in machine learning related development.

Textbook and References

Primary references

A set of lecture slides will be provided to you which will serve as the primary reference for our course. The content of these slides will be drawn from the following textbooks:

- "Learning from Data" by Lin, Magdon-Ismail, and Abu-Mostafa
- "Mathematics for Machine Learning" by Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong
- "Deep Learning with Python (2nd Ed)" by François Chollet

Recommended supplementary references

The following closely-related and well-written textbooks will serve as supplementary materials for us. You can consult these textbooks to gain a different perspective or to develop your project ideas.

- "Understanding Machine Learning: From Theory to Algorithms" by Shalev-Shwartz and Ben-David
- "Pattern Recognition and Machine Learning" by Christopher Bishop
- "Statistical Pattern Recognition (3rd ed.)" by Webb and Copsey
- "The Elements of Statistical Learning (2nd ed.)" by Friedman, Tibshirani, and Hastie
- "Machine Learning: An Algorithmic Perspective (2nd ed.)" by Stephen Marshland
- "Foundations of Machine Learning" by Mehryar Mohri
- "Introduction to Machine Learning" by Ethem Alpadin
- "Machine Learning: A Probabilistic Perspective" by Kevin Murphy
- "Reinforcement Learning (2nd ed.)" by Sutton and Barto

Programming references

- "Hands-On Machine Learning with Scikit-Learn and TensorFlow" by Aurélien Géron
- "Practical Programming, An Introduction to Computer Science Using Python" by Montojo et al.

Optional readings

There are optional readings for each week of the class. These are papers written by some of the key contemporary thinkers on the topic of machine learning. **You will not be graded based on these optional readings.** They are meant for you to gain a broad perspective on the topic of this course, as well as for you to practice reading academic publications, which is a crucial skill whether you end up in academia or industry. Tip: read the introduction and the experimental section first. If the paper interests you, then read it more in depth!

Grade Breakdown

You will be evaluated based on three components: homework, term tests, and a final team-based project. The grade breakdown for each component as is follows:

Homework 20% **Term Test I** 20% **Term Test II** 20% **Final Project** 40%

Homeworks (20%)

There will be a homework assignment roughly every two weeks. A subset of the problems will be randomly drawn and graded based on the clarity and (less so) on correctness of your answers. These assignments will often include several short programming questions. Programming questions should be done in Python using a ML library called Skorch, which combines two other popularly used libraries, Scikit-learn and Pytorch. **Homeworks should be printed out and handed in at the beginning of the class with your name and student ID clearly labeled.**

Term Tests (Test I 20% + Test II 20%)

There are two in-class term tests, which will be closely modeled after your homework assignments. They will consists of computational questions that assess your understanding of the material and ability to extrapolate what you've learned to new scenarios. Calculators are not required. These tests will take place during regular class hours. The tentative schedule for these tests are set to occur during Week 7 and Week 13 or 14 (dates are to be announced on Canvas).

Final Project (Proposal 5% + Presentation 5% + Report 30%)

A major component of this course involves a final project. Each project will be performed by a team of $1-3$ students. The projects should fall under one of the following categories:

- **Solving a problem involving machine learning as a sub-component** (e.g., translating foreign language content through a smart phone camera)
- **A survey that deeply explains an advanced area of application and a collection of techniques associated with the application** (e.g., machine learning techniques for generating cartoon videos)
- **Independently proposed topic** (e.g., comparison between a group of ML techniques for solving a certain problem, project related to your team's research or academic interests, theoretical-papers are welcome)

All types of project must involve one or more simulation component and utilization of at least one concept that we have discussed in class. There is no requirement on the programming language for the final project.

Sign-up and team formation

A sign-up sheet will be made available to you with a list of pre-determined potential projects. You are welcome to choose and modify upon the projects or propose your own projects. You should have formed a team by the end of Week 2 (9/1, 11:59 pm). If you have not found a team by then, contact me. Each team will be issued a team number. Your team will have the option to choose a team name.

Project proposal (5%)

A project proposal must be **submitted through Canvas** by the end of Week 5 (9/22, 11:59 pm). The format of the project proposal is as follows:

- A maximum of 2 pages. Clearly pace your tentative project title, team number, (optionally) team name, names of all team members and student IDs at the top of the page. LaTex is strongly recommended for generating your project proposal.
- Page 1 should contain one or two short paragraphs describing your problem, a justification for its relevance, and your preliminary ideas on the next steps.
- Page 2 should contain a description of the potential cost involved, project timeline, responsibility of team members along with any relevant diagrams/figures. This page is optional.

You are recommended to start working on your project as soon as you have an idea, but I might communicate with your team regarding the feasibility of the project and other considerations after receiving the proposal from your team. Failing to submit the proposal will result in 0%.

Project presentation (5%)

A presentation of your project will occur at the last day of the class during Week 14. You and your team will describe to the rest of the class the problem you are trying to solve and the solution that your team had came up with. The presentation style should be kept conversational and informal. **The project presentation is graded based entirely on the clarity of your presentation**. Grade is deducted for the whole team if you or your team members cannot adequately explain one or more concept/idea/equation on the presentation slide.

Final report (30%)

The format of the final report (maximum 1 + 8 + 2 + 1 = 12 pages) is as follows:

- 1 cover page indicating the title of the project, your team number (and name), names of all students along with their student IDs and a rough estimate of the percentage of contribution from each student in terms of the overall report. For example, if all three students contributed equally, then the percentage is 33%.
- You are allowed a maximum of 8 pages of main report (you will not be penalized for having less than 8 pages). The main report should contain an introduction, a related work section, a presentation of your main results (and simulation) and a conclusion or future work.

- You are allowed a maximum of 2 pages of appendices (additional figures, terminologies, data, tables) and a maximum of 1 pages of citations (citation format must be consistent).
- The report should be typed in 12 point font. LaTex is strongly recommended for generating your final PDF report.
- The report **must** contain a diagram of the overall system showing all inputs, outputs and data/signals flowing between internal components.
- The appendix should contain a list terminologies used in the report. If this is not included, then all terminologies should be clear from the main report.
- Your team will pick the best template for presenting your material. The margin and spacing must be be legible. An unreadable report will receive a grade of 0%.

The final report will be graded out of 100 points based on the following criteria:

- Is the problem you are trying to solve clearly defined? [10 pts]
- Is the problem you are trying to solve relatively under-explored or are there already well-known solutions to your problem? [10 pts]
- Does your report follow the report format guidelines? [10 pts]
- Does your report contain a diagram of your ML system showing all input and outputs as well as data/signal between each sub-component? If so, is it clear to the readers? [10 pts]
- Is your report well-organized? Does your report make sense when read (does each word, sentence and paragraph make sense)? Is it mostly free of typos and other serious grammatical mistakes? Does it flow logically? Are there sufficient citations? [20 pts]
- Is your report technically sound? Does the experiment/simulation make sense? Are there sufficient mathematical justification or citation for technical claims? Is it mostly free of mathematical mistakes? Are equations clear? Are most mathematical symbols defined? [20 pts]
- Does your report clearly define most concepts and technical terminologies? Is the report self-contained? Are all figures clear and clearly labeled? [20 pts]

A rule-of-thumb is: every time something confusing or unclear appears in the report, a point may be deducted. The final report is to be submitted through Canvas approximately one week after our last class.

Your Final Grade

Final grades at Georgia Tech are entered as **whole letter** grades (**A, B, C, D, F**). These grades correspond to the following percentage:

A 90-100% **B 80-89%** **C 70-79%** **D 60-69%** **F 0-59%**

If you have performed satisfactorily in this course but was unable to meet the full requirements of the course due to nonacademic circumstances beyond your control, contact me to see if you are eligible for a grade of I (Incomplete).

Tentative Course Schedule

The schedule for each week is shown in the following table.[1] The topics are subject to change based on scheduling as well as the interests or need of the class.

Week	Lecture Content	Reference	Task
Week 1 (8/21-8/25)	Machine Learning Problems, Perceptron, K Nearest Neighbor	LFD 1.1 - 1.2, 3.1, 6.1-6.2.4; MML 8.1	A1 Released!
Optional Reading	• "AI as a control problem" by R. Sutton • "Early History of Machine Learning" by A. Fradkov, 2020 • "Machine learning: Trends, perspectives, and prospects" by M. I. Jordan & T. M. Michell, 2015		
	(Shallow) Supervised Learning		
Week 2 (8/28-9/1)	Linear and Logistic Regression, Feature Normalization, Regularization	LFD 3.2-3.3, MML 8.2, 9.1 - 9.2	
Optional Reading	• "Weight Decay Can Improve Generalization" by A. Krogh and J. A. Hertz		
Week 3 (9/4-9/8)	Softmax Regression and Multilayer Perceptron	LFD 7.1 - 7.6	A1 Due, A2 Released!
Optional Reading	• "Learning representations by back-propagating errors" by Rumelhart, Hinton, Williams • "A critique of pure learning and what artificial neural networks can learn from animal brains" by Zador		
	(Deep) Supervised Learning		
Week 4 (9/11-9/15)	Convolutional Neural Network (CNN), Residual Network, Adversarial Attack, Interpretability, Distillation	DLPy 8.1 - 8.3	
Optional Reading	• "Neocognitron" by K. Fukishima • "ImageNet Classification with Deep Convolutional Neural Networks" by Krizhevsky, Sutskever, Hinton		
Week 5 (9/18-9/22)	Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), Long-Short Term Memory (LSTM)	DLPy 10.1 - 10.4	A2 Due, A3 Released!

[1] LFD = Learning from Data, MML = Mathematics of Machine learning., DLPy = Deep Learning with Python

Week	Lecture Content	Reading	Task
Week 6 (9/25-9/28)	Support Vector Machines, Kernels	LFD 8.1 - 8.4; MML 12.1 - 12.6	
Optional Reading	• "LightGBM: A Highly Efficient Gradient Boosting Decision Tree" by Ke et al., 2017		
9/29-10/6	Mid-Autumn Festival and National Day No Class		
	Learning Theory		
Week 7 (10/8-10/13)	Generalization Bound	LFD 1.1 - 1.4.2	A3 Due, A4 Released!
Optional Reading	• "Model complexity of deep learning: a survey" by Hu et al., 2021		**Term test 1 [10/8]**
Week 8 (10/16-10/20)	VC dimension, Bias and Variance	LFD 2.1 - 2.3.2	
Optional Reading	• "Local Rademacher complexities" by Bartlett et al., 2005		
	Dimensionality Reduction		
Week 9 (10/23-10/27)	Principal component analysis, Autoencoder	MML 10.1 - 10.8, DLPy 12.4.3	A4 Due, A5 Released!
Optional Reading	• "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection" by Belhumer et al., 1997		
	Unsupervised Learning and Generative Models		
Week 10 (10/30-11/3)	K-Means, Gaussian Mixture Model (GMM), Expectation-Maximization Algorithm	LFD 6.3.3 - 6.3.4, MML, 11.1 - 11.5	
Optional Reading	• "Gaussian Mixture Models for Parking Demand Data" by Fiez and Ratlliff		
Week 11 (11/6-11/10)	Variational Autoencoder, Generative Adversarial Network (GAN)	DLPy 12.4 - 12.5.7	A5 Due, A6 Released!
Optional Reading	• "Artificial Curiosity & Creativity Since 1990-91" by J. Schmidhuber		
	Reinforcement Learning (RL)		
Week 12 (11/13-11/17)	Online Learning, Learning in Games	See Notes	
Optional Reading	• "Multiagent learning using a variable learning rate" by Bowling and Veloso, 2002		
Week 13 (11/20-11/24)	MDP-based RL, Policy Gradient, Imitation Learning	See Notes	A6 Due!
Optional Reading	• "Machine Learning Systems are Stuck in a Rut" by Barham and Isard		**Term test 2 [11/24]**
Week 14 (11/27-12/1)	Project Presentation		**Final report due 12/9 11:59 pm via Canvas**

Important Administrative dates

- August 21: First day of classes
- August 25, by 4:00pm ET: Deadline for registration and class schedule changes
- August 25: Deadline for Submitting Online Application for Graduation (OAG) for fall graduation
- August 28, by 4:00pm ET: Deadline to pay tuition & fees
- September 29 – October 6: Mid-Autumn Festival & Chinese National Day holiday break
- October 28, by 11:59pm ET: Deadline for dropping a course with "W" grade
- October 28: Deadline to change grade mode (pass/fail/letter grade)
- December 7 - 14: Final exams for on-site classes/ video classes from GT-Atlanta
- December 19, 6:00pm ET: Grades are available

Additional Policies

Academic integrity

Any student suspected of cheating or plagiarizing on a test or homework will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations. Redistributing materials from this course (e.g., contributing to test banks, CourseHero, Chegg, or similar sites) is prohibited. For additional information, please visit: `https://policylibrary.gatech.edu/student-life/academic-honor-code`.

Collaboration and group work

Students are strongly encouraged to discuss homework problems with each other. However, each student must write up and turn in their own solutions. Suspicious homework submissions will be immediately be referred to the Office of Student Integrity.

Absences, late homeworks, and missed tests

You will not be penalized for any excused absences (such as: illnesses, religious observances, career fairs, job interviews). I cannot accept late homeworks in the absence of prior approval. If you expect to miss an exam, please contact me as soon as you realize this so we can make alternative arrangements. We may consider options to take the exam at an alternate time or instead may adjust the grading allocation to place more emphasis on other exams, depending on the circumstances.

Re-grading policy

Re-grading request should be submitted via email to the instructor. Your letter must include student name, student ID and a justification for the request, which refers specifically to the student's answers and course material. Requests without this justification will not be considered. The deadline for requesting a re-grade is one week after the grades are returned. Re-grade may result in a decrease of the grade.

Accommodations

If you are a student with learning needs that require special accommodation (e.g., disability, distress), contact the me as soon as possible in order to make an appointment to discuss your special needs and to obtain an accommodations letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs.

Student-faculty expectations

Respect for knowledge, hard work, and cordial interactions will help build a positive environment that is conducive to our learning. See `www.catalog.gatech.edu/rules/22` for some basic expectation that you can have of us and that we have of you.