Math 156 — Machine Learning — Summer Session A

The class will consist of lectures with me (the instructor), discussion section with the TAs, and optional office hours all of which will occur on Zoom and/or video recordings. All times listed throughout are in PST.

Course websites: CCLE: https://ccle.ucla.edu/course/view/211A-MATH156-1

Lecture time and location: MWR 09:00–10:50, virtual on Zoom

Instructor: Dr. Dominique Zosso

E-mail: zosso@math.ucla.edu
Howto: www.insidehighered.com/views/2015/04/16/advice-students-so-they-dont-sound-silly-emails-essay

Virtual office hours: MWR 12:30–1:30pm
Zoom room for class and office hours: PMI = 377 051 7025 / Passcode = zXvH38
Link: https://zoom.us/j/3770517025?pwd=VzI2YUFLdTNHdjdFRHo5bWlrRUkvQT09

Section 1A: T 9:00–10:50am

Teaching Assistant: Jiawei Zhang

E-mail: jwzcharlie@gmail.com

Virtual office hours: MW 2:00–3:00pm
Zoom-room for discussion section and virtual office hours: PMI = 980 3627 6047 / Passcode = 161218
Link: https://ucla.zoom.us/j/98036276047?pwd=RlpmNW1lUGE4ZzRqT1I2Z3l6dlQ3dz09

Section 1B: T 11:00am–12:50pm

Teaching Assistant: Zeyu Wang

E-mail: zeyuwang@g.ucla.edu

Virtual office hours: RF 5:30–6:30pm
Zoom-room for discussion section and virtual office hours: PMI = 803 045 2681, no passcode
Link: https://ucla.zoom.us/j/8030452681

Course text: *Pattern Recognition and Machine Learning*, by Christopher M. Bishop, Springer, 2006 (ISBN-13: 978-0387-31073-2), *plus complementary sources provided where necessary*. The book is freely available as pdf from the author:
https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf

Description: "Machine Learning" is an *introductory course* on mathematical models for pattern recognition and machine learning. The students will become familiar with fundamental concepts such as learning of parametric and non-parametric probability distributions, the curse of dimensionality, correlation analysis and dimensionality reduction, and concepts of decision theory. Advanced machine learning and pattern recognition problems will be covered, including data classification and clustering, regression, kernel methods, artificial neural networks, and Markov-based models such as hidden Markov models and Markov random fields. While these methods are fairly generic and widely applicable, they will be accompanied and illustrated by practical examples drawn from imaging, computer vision, document and social network analysis, etc.

Learning outcomes: Upon completion of the course, students will be able to:

1. Describe and understand the mathematics of basic models used in machine learning, and their training
2. Explain various mathematical approaches to dimensionality reduction with PCA (minimum error, maximum variance, probabilistic)
3. Understand the mathematical underpinnings of linear models for regression and classification, and kernel-based extensions
4. Understand and apply basic artificial neural network structure and training, from perceptron to multilayer networks
5. Build, train, and use basic graphical models such as Hidden Markov models (fields and chains)

Prerequisites: courses 115A, 164, 170A or 170E or Statistics 100A, and Computer Science 31 or Program in Computing 10A. Strongly recommended requisite: Program in Computing 16A or Statistics 21.

Computing device: It is important that students be able to use a personal computing device for this class, for completion of some practical classwork. This can technically be any device with a web-browser to run MATLAB online (see below; note that Safari on iOS devices is not optimal).

Homework: Homework problems will be assigned regularly. Homework problems are integrative part of the curriculum. It is strongly encouraged to routinely check any paper-and-pencil calculations with MATLAB (Emphasis on: as a check, not to get the solution in the first place). Homework assignments will not be collected, corrected or graded. But, to learn this material (and to do well on exams and quizzes), you should master all of the homework problems. Indeed:

Quizzes: About 2 times a week, there will be a quiz roughly covering current lectures and homework problems, available as PDF on the course website and due 24 hours later, on gradescope. The quizzes will take you 5–10 minutes to complete, but are not timed. The lowest (or a missed) quiz will be dropped, but make-up quizzes will not be given.

MATLAB assignments: 8 MATLAB-based lab assignments with practical problems will be posted on the course website. Lab code and reports are to be submitted as pdf-files through gradescope, and will be graded. The lowest (or a missed) lab will be dropped.

Exams: There will be a midterm, due on **July 8**, as 24-hour takehome exam (it should take you about 60–75 minutes to complete). There will be a final exam due on **July 29**, as 24-hour takehome exam (it should take you about 90–120 minutes to complete). The final exam will cover the entire semester, but emphasize later parts. Failure to take the final exam will result in an automatic fail grade being assigned. **All exams are open book: your books and notes are permitted, however, NO other resources (calculators, online tools, friends, etc.) will be allowed during exams.** All exams will become available on the course website as PDF 24 hours before the due date/time: exams will be due at noon the day marked on the schedule, and become available no later than at noon the day before (i.e., the midterm becomes available at noon of W, 7/7, and will be due by R, 7/8, at noon, PST).

To take an exam, you will be responsible for

- either: printing the exam, filling it out by hand, scanning your copy, and submitting the resulting pdf to gradescope,
- or: downloading the pdf, completing it electronically, and submitting the edited pdf to gradescope.

Grading policy: 7 (out of 8) Quizzes: 20%, 7 (out of 8) Labs: 20%, 1 Midterm: 30%, Final: 30%. All scores will be available on gradescope. The instructor will translate scores into letter grades according to his best judgment and Department policy.

Grading Scales: The lab assignments in this course will be graded on the following simplified scale:

4 — excellent work; no real complaints on content, code commenting or on writing.

3 — work basically correct but missing some details/less clearly argued or commented than I would like.

2 — argument mostly correct, but there is a misstep in the mathematics, choice of algorithm. An especially poorly written answer, paper or commented code might merit a 2, as well.

1 — serious gaps in the mathematics, some ideas in the right direction, but didn't really get anywhere.

0 — didn't do the problem or it was completely wrong.

Getting and Using MATLAB

Site license: UCLA has a campus-wide license for MATLAB:
see `https://softwarecentral.ucla.edu/matlab-getmatlab` for more information/instructions.

If you do not want to install MATLAB, locally (it is quite big), you may use "Apporto"-virtual machines (be in touch with support@pic.ucla.edu for help on that). Most conveniently, though, we suggest using MATLAB on the web:

MATLAB on the web: To avoid the need of installation, we suggest using MATLAB on the web, instead. This service is available to students, staff, and faculty of UCLA at no cost. The only requirement is to sign up for a MATLAB account with a `ucla.edu`-email-address.

Follow the instructions on `https://www.mathworks.com/academia/tah-portal/ucla-31454052.html` to create your UCLA-MATLAB account. Once complete, you have access to MATLAB online at `matlab.mathworks.com` (iOS safari browser is discouraged). Note that your MATLAB workspace is persistent across sessions, devices, and there is an online filesystem!

MATLAB Reports: Your solutions to the lab assignments should include a clear write up of your solutions (LaTeX is strongly preferred, but other text editors or handwritten form is fine as well) including an English description of your solution to the programming exercises (explain your process and the high level action of your code). Reports should include answers to the specific questions/tasks set forth in the assignment. Please also include a discussion and reflection of what the practical experiments reveal, to you. For each MATLAB assignment, you will need to submit a single .pdf-file to gradescope with your code, experiments, figures, results, and observations.

`.m`-files or Livescripts: You are encouraged to organize your online filesystem appropriately. It is recommended that for each MATLAB assignment, you create one or more `.m`-files or livescripts as required. Make sure to include an appropriate amount of comments to make your code understandable. Use `disp(...)` to print values of variables on the screen, and use plots and figures as suitable. You can create sections to your script (for reporting purposes) by using `%%`-signs. Use comments at the end of the file to discuss results and observations as required.

publish and submit: A simple way is to use MATLAB's **publish**-function to create .pdf-reports. The publish button will run the current file, and render the code along with output (command line and figures) into a single .pdf-file. **This method is simple enough—it is your responsibility though to make sure the report is complete!** Note: The publish function *does in general not play well with functions that require parameters, and scripts that include user interaction* (e.g. `input`). Use a wrapper script that calls such a function with parameters, and include the code of the inner function using `<include>`.

A few helpful items[1]:

- Use `%% ...` to create a new section; this embeds all results of a section before moving on to the next.
- Use `% comments...` to include your observations.
- To include the code of *separate* function files, use `<include>` *exactly* as follows (extra comment lines and spacing between `%` and `<`, etc., matters!):

```
%% Beginning of the wrapper script which we will publish
%
% <include>SeparateFunctionFileToBeIncludedInTheReport.m</include>
%
```

LaTeX reports: A possibly preferable alternative to MATLAB's publish feature is to use LaTeX to properly typeset your report, including background, comments, code, experimental setup, results, discussion, and conclusion.

MATLAB due dates: Either way, submit your report as a single pdf-file per assignment, through gradescope. Your reports for the 8 labs will be due on Fridays, each week (see schedule, below), by 6:00pm PST.

[1] Check `https://www.mathworks.com/help/matlab/matlab_prog/marking-up-matlab-comments-for-publishing.html`

Tentative class schedule:		Section numbers (§) refer to the book; topic numbers refer to lecture notes.	
M 6/21	§1.2, §1.5, §1.6	1. Course introduction. Recap on linear algebra, probabilities	
	§2.3, §2.4	2. Gaussian, exponential pdf; Learning parametric pdf	
T 6/22		Discussion: HW 1, Lab 1	
W 6/23	§2.5	3. Learning non-parametric pdf	**Quiz 1 due**
R 6/24	§12.1	4. PCA: maximum variance, minimum error, high-dimensional PCA	
	§12.2	5. Probabilistic PCA (ML-PCA, EM, Bayesian PCA)	**Quiz 2 due**
F 6/25			**Lab 1 due**
M 6/28		**JUNETEENTH HOLIDAY OBSERVED** (no class)	
T 6/29		Discussion: HW 2, Lab 2 & 3	
W 6/30	§12.3	6. Non-linear latent variable models: kPCA	
	§3.1	7. Linear basis function models, least squares and maximum likelihood	**Quiz 3 due**
R 7/1	§3.3	8. Bayesian linear regression	
F 7/2			**Labs 2 & 3 due**
M 7/5		**INDEPENDENCE DAY HOLIDAY OBSERVED** (no class)	
T 7/6		Discussion: HW 3, Lab 4	
W 7/7	§4.1	10. Discriminant functions; least squares	**Quiz 4 due**
	§4.2, §4.3	11. Logistic regression: prob. generative & discriminative models	
R 7/8		**MIDTERM EXAM** (24 hour takehome)	
F 7/9			**Lab 4 due**
M 7/12	§14.2, §14.3	12. Mixture of linear classifiers: Boosting and Bagging	
T 7/13		Discussion: HW 4, Lab 5	
W 7/14	§9.1	13. k-Means	
	§9.2, §9.3	14. Gaussian mixture model, Expectation-Maximization	**Quiz 5 due**
R 7/15	§6.1, §6.2	16. Dual representation, kernel trick; Constructing kernels	
	§6.4	17. Gaussian processes, GP regression, GP classification	
F 7/16			**Lab 5 due**
M 7/19	§7.1	18. Support vector machines, k-SVM	
T 7/20		Discussion: HW 5, Lab 6 & 7	
W 7/21	§4.1.7, §5.1	19. Biological motivation; The perceptron; Feed-forward Network	
	§5.2, §5.3	20. Network training	**Quiz 6 due**
R 7/22	§8.1	21. Bayesian Networks	
	§8.3	22. Markov Random Fields; Iterated conditional modes	**Quiz 7 due**
F 7/23			**Lab 6 & 7 due**
M 7/26	§13.1, §13.2	23. Hidden Markov Models, forward-backward, Viterbi algorithm	
T 7/27		Discussion: HW 6, Lab 8	
W 7/28	n/a	15. Spectral clustering	
		Leeway and Review	**Quiz 8 due**
R 7/29		**FINAL EXAM** (24 hour takehome)	
F 7/30			**Lab 8 due**

Behavioral expectations:

You are expected to follow the UCLA Student Conduct Code, which can be found here: https://www.deanofstudents.ucla.edu/Individual-Student-Code. If you are not familiar with it, take a few minutes to read about your responsibilities (and mine!). Violations will be referred to the Dean of Students.

From the office of the Dean of Students: "With its status as a world-class research institution, it is critical that the University uphold the highest standards of integrity both inside and outside the classroom. As a student and member of the UCLA community, you are expected the demonstrate integrity in all of your academic endeavors. Accordingly, when accusations of academic dishonesty occur, The Office of the Dean of Students is charged with investigating and adjudicating suspected violations. Academic dishonesty includes, but is not limited to, cheating, fabrication, plagiarism, multiple submissions or facilitating academic misconduct." Students are expected to be aware of the University policy on academic integrity in the UCLA Student Conduct Code: https://www.deanofstudents.ucla.edu/Portals/16/Documents/UCLACodeOfConduct_Rev030416.pdf. Please note the sections on (1) cheating, (2) plagiarism, and (3) unauthorized study aids.

Zoom etiquette: Remote learning is hard. I am going to do my best to give you many resources (most importantly time) to gain a good experience from this course. That said, I know that many things will be lacking from this experience and parts of it may be overwhelming. I welcome feedback (immediate and after the course ends) regarding the course, material, and set-up. I hope we will find ways to have good connection during these six weeks, but I know that the electronic communication tools can be a struggle.

Below are a list of some things to keep in mind with use of these tools.

- I hope you will consider sharing your video when on Zoom (if you can), but always feel free to turn off your video and audio.
- You can always feel free to participate in the course asynchronously (e.g., watch- ing videos of lectures or reviewing notes PDFs rather than attending lecture on Zoom).
- Large group meetings on Zoom are a challenge and will require extra effort to communicate politely. In general, please keep your microphone muted when you are not speaking and if you expect there will be many looking to speak, use the "raise hand" or chat feature. **Since I might easily miss that there is a chat message or hand raised, please don't hesitate to speak up and alert me to this fact!**

Zoom recording: The instructor, the TAs might/will record the lecture/discussion Zoom meetings. Enrolled students and participants of the course will receive advance notice of any such recording and can always opt out of video/audio/chat participation. Students are not allowed to record any part of the lecture or discussion sessions, including video, audio, chat messages, and shared contents.

Notice about sexual harassment, discrimination, and assault: Title IX prohibits gender discrimination, including sexual harassment, domestic and dating violence, sexual assault, and stalking. Students who have experienced sexual harassment or sexual violence can receive confidential support and advocacy from a CARE advocate:

The CARE Advocacy Office for Sexual and Gender-Based Violence
1st Floor, Wooden Center West
CAREadvocate@caps.ucla.edu
(310) 206 – 2465

You can also report sexual violence or sexual harassment directly to the University's Title IX Coordinator:

Kathleen Salvaty
2241 Murphy Hall
titleix@conet.ucla.edu
(310) 206 – 3417