AI in Genetics
ZOO6927 / BOT6935 / ZOO4926
Fall 2024
Syllabus

Class Number *29890 / 29408 / 29411*

Class Meetings

Tuesday | (3:00 PM - 4:55 PM) Room: FAC127
Thursday | (3:00 PM - 3:50 PM) Room: FAC127

Instructors

Juannan Zhou, Assistant Professor
Department of Biology
E-mail: juannanzhou@ufl.edu
Office: Bartram 122
Office Hours: Thursday 4:00-5:00 PM

Pre-requisites

There are no formal prerequisites. The course assumes familiarity with basic transmission genetics ("Mendelian genetics"), molecular genetics, basic algebra, and probability. Some knowledge of linear algebra and calculus will be useful but not assumed.

Course Objectives

The overall objective of this course is to provide a comprehensive overview of applications of modern machine learning techniques in various areas of genetics. I aim to use this course to provide opportunities for students to (1) integrate machine learning into their own research; and (2) learn critical computational and statistical skills that will hopefully broaden the student's career path. The course will cover applications of AI to genomics, gene expression and regulation, protein design and evolution, molecular evolution, population, and quantitative genetics.

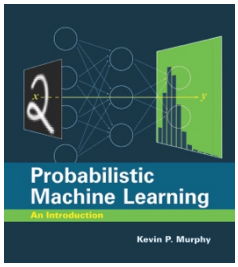
Objectives of the course will be achieved if, by its conclusion, students can:

- Understand the basic concepts and mathematical/statistical theory behind modern machine learning methods
- Understand 80% of most research papers applying machine learning methods in fields relevant to the student's own research, and grasp the basic ideas of technical machine learning papers
- Develop new research questions that are well-suited for applying machine learning methods to improve their current studies; or identify existing questions where machine learning offers a potentially superior alternative to the current approaches.
- Identify the right machine learning frameworks and tools for answering these questions
- Build machine learning models to solve specific questions using coding languages such as Python
- Get the model to work by using different model architectures and training methods

Course format

The course will begin with a concise introduction to the mathematical and statistical foundations of modern machine learning. We will then learn in detail the most popular machine learning models, including MLP, CNNs, transformers, and generative models. The rest of the course will focus on student-led discussions of a selection of recent papers covering different subjects of AI in genetics.

Textbook & Online Resources



A. Textbook (recommended)

Probabilistic Machine Learning: An Introduction by Kevin Patrick Murphy. MIT Press, March 2022.
Free pdf: https://github.com/probml/pml-book/releases/latest/download/book1.pdf

B. Course Website (Canvas)

Class material including the syllabus, discussion readings, and problem sets, exam results, some lecture slides and other information related to the course will be posted on the course Canvas website (http://elearning.ufl.edu). You are responsible for **all** announcements made in lecture and/or posted on the course website for this class. For help with Canvas, call the UF Computing Help Desk at 352-392-4357, or visit the Canvas support website: https://elearning.ufl.edu/keep-learning/.

Assessments and Grading

a. Exam

The course will have one mid-term exam. The grade of the mid-term exam will account for **40%** of the student's final grade. Format of the exam will be take-home and consists of of practical questions the student needs to solve using their preferred coding language.

b. Final project

Each student is expected to complete a final project, which will make up **50%** of the student's final grade. For the project, the student should apply machine learning techniques to solve a biological problem, preferably directly relevant to their thesis work.
The student will present their results in class and turn in a term paper. The paper should be similar in format to a conference publication (e.g. https://proceedings.mlr.press/v240/).

c. Class Participation.

Each student will be responsible for leading an in-class discussion on one of the assigned readings. This will account for **10%** of final grade.

d. Grading Scale

Point Range (%)	Letter Grade	GPA equivalent
≥ 90.00	A	4.0
86.7 – 89.9	A-	3.67
83.3 – 86.6	B+	3.33
80.0 – 83.2	B	3.0
76.7 – 79.9	B-	2.67
73.3 – 76.6	C+	2.33
70.0 – 73.2	C	2.0
66.7 – 69.9	C-	1.67
63.3 – 66.6	D+	1.33
60.0 –63.2	D	1.0
56.7 – 59.9	D-	0.67
< 56.7	E	0

e. Grading policies

Grade Curve Policy

The grades shown in the table are guarantees, e.g., if you make AT LEAST a 73.3% you are GUARANTEED a C+. I reserve the right to curve downward, i.e., to be more generous.

Make-up Exam Policy

Make-up exams will be administered on a case-by-case basis. Valid excuses include (but are not necessarily limited to) personal illness or injury or the illness, injury, or death of a family member. If you know you will need to miss class (e.g., for a job interview), please notify me in advance.

Honesty Policy

- All students registered at the University of Florida have agreed to comply with the following statement: "I understand that the University of Florida expects its students to be honest in all their academic work. I agree to adhere to this commitment to academic honesty and understand that my failure to comply with this commitment may result in disciplinary action up to and including expulsion from the University."
- In addition, on all work submitted for credit the following pledge is either required or implied: "*On my honor I have neither given nor received unauthorized aid in doing this assignment*."
- If you witness any instances of academic dishonesty in this class, please notify the instructor or contact the Student Honor Court (392-1631) or Cheating Hotline (392-6999). For additional information on Academic Honesty, please refer to the University of Florida Academic Honesty Guidelines at: http://www.dso.ufl.edu/judicial/procedures/academicguide.html. **Accommodation for Students with Disabilities**

- Students who will require a classroom accommodation for a disability must contact the Dean of Students Office of Disability Resources, in Peabody 202 (phone: 352-392-1261). Please see the University of Florida Disability Resources website for more information at: http://www.dso.ufl.edu/drp/services/.
- It is the policy of the University of Florida that the student, not the instructor, is responsible for arranging accommodations when needed.

Once notification is complete, the Dean of Students Office of Disability Resources will work with the instructor to accommodate the student.

Software Use

All faculty, staff and student of the University are required and expected to obey the laws and legal agreements governing software use. Failure to do so can lead to monetary damages and/or criminal penalties for the individual violator. Because such violations are also against University policies and rules, disciplinary action will be taken as appropriate.

Class schedule

The course schedule is divided into two units. Unit 1 covers the mathematical foundation and basic concepts in modern machine learning. This unit is consisted of lecture-led presentations and discussions. Unit 2 focuses on various special topics in ML applications to genetics and consists of **student-led discussion** of research papers. Please note that the class schedule is subject to adjustment by the lecturer according to class size and course progression. Student nomination of special topics and papers are welcome.

Date	Week	Subject	Topics
8/22/24	0	Course introduction; Mathematical foundation	Linear algebra
8/27/24	1	Mathematical foundation	Linear algebra
8/29/24	1	Mathematical foundation	Probability
9/3/24	2	Mathematical foundation	Probability
9/5/24	2	Machine learning basics	Multi linear perceptron; Backprop; Autodiff; Gradient descent
9/10/24	3	Machine learning basics	Training neural networks; Regularization
9/12/24	3	Machine learning basics	Convolutional neural networks
9/17/24	4	Machine learning basics	Language models; RNNs; Transformers
9/19/24	4	Machine learning basics	Graphical neural networks; Generative models
9/24/24	5	Machine learning basics	Generative models
9/26/24	5	Machine learning basics	Non-parametric methods; Gaussian processes
10/1/24	6	Paper discussion	Bioinformatics
10/3/24	6	Paper discussion	Proteins
10/8/24	7	Paper discussion	Proteins
10/10/24	7	Paper discussion	Gene expression and regulation
10/15/24	8	Paper discussion	Gene expression and regulation
10/17/24	8	Paper discussion	Genomics
10/22/24	9	Paper discussion	Genomics
10/24/24	9	Paper discussion	Molecular Evolution
10/29/24	10	Paper discussion	Molecular Evolution
10/31/24	10	Paper discussion	Population Genetics
11/5/24	11	Paper discussion	Population Genetics
11/7/24	11	Paper discussion	Quantitative Genetics; Plant/Animal Breeding
11/12/24	12	Paper discussion	Quantitative Genetics; Human diseases
11/14/24	12	Paper discussion	Generative models in genetics
11/19/24	13	Student presentations	
11/21/24	13	Student presentations	
11/26/24	14	Thanksgiving break	
11/28/24	14	Thanksgiving break	
12/3/24	15	Student presentations	