# COURSE DESCRIPTION AND POLICIES

Please read this **3-page document** carefully. If you have any questions, contact the instructor.

This is a basic course in Machine Learning covering the theory, algorithms, and applications. The prerequisites are Ma 2 and CS 2, or equivalent (familiarity with basic probability, basic matrices, and basic computer programming). The course is given in "flipped classroom" mode, where recorded lectures are used for instruction and class time is focused on discussion of the lecture contents.

## Class website

The only source for homework, final exam, and solutions for this course is the class Canvas website and what it links to.

## https://caltech.instructure.com/courses/6282

All other links that you need are included in that website. Please do not look at homework, final exam, or solutions from any other source or website (not even other Caltech sites). This is a strict (honor code) rule.

#### Textbook

There is no required textbook for the course. The recommended textbook is:

Y. Abu-Mostafa et al, Learning From Data - a Short Course, AMLbook.com, 2012

which covers 14 out of the 18 lectures and is available at Amazon. The rest is covered by e-chapters of the book that are available to you for free at:

## https://amlbook.com/eChapters.html

Password given in Canvas; please do not share.

# **Office Hours**

The course personnel and office hours are listed on the class website. The TA office hours are scheduled close to the day when the homework is due, and are held in 139 Moore which has direct access from the north after hours.

## Piazza

You can ask questions and discuss any items in person with the instructor or the TA's during their office hours. Other than that, the only online mode of asking *technical questions about homework problems* will be through the Piazza discussion forum (linked from Canvas), not by email. You have the option in Piazza to ask questions anonymously if you so choose. The instructor and TA's will answer questions on Piazza on a regular basis. Administrative questions of private nature (e.g., discussing a homework grade) should be done by email.

#### Homework and Final

Eight weekly homework sets with 10 problems each and a final exam with 20 problems (each problem carries 10 points) will be made available at a specified schedule of dates and times on the class website. The schedule is strict with no extensions. Please turn in your homework in Gradescope (linked from Canvas).

## Not Just Multiple Choice

The homework and final exam are formatted as multiple-choice problems, but your answer must be detailed. In particular,

1. Start each problem with a full derivation of your answer. If it's an experiment, include the code you wrote and report the numerical answer that it got you. If it is a derivation, carry it out to get the answer. If it is analytic reasoning between the multiple choices, argue for each choice why it is valid or not valid independently of the other choices. This derivation carries 6 out of the 10 points for each problem. There will be partial credit if your derivation has errors.

**2.** You cannot justify your answer by the fact that the correct answer must be one of the multiple choices.

3. Your derivation must be your own; you cannot copy from any source.

4. You can use any programming language or platform, including macros and packages, but don't use other people's code that was written for this class.

5. After you write your derivation, choose the matching multiple-choice answer. This carries 4 points out of the 10 points. If you make the right choice, you get 4 points. If not, you lose the points. You have the option of "hedging" by reporting two of the choices as possibilities. In that case, if one of them is right you get 2 points only, and if neither is right you lose the points.

### **Course Grade**

The course grade will be generally based on the total score of the 8 homework sets and the final. Your worst homework will be discarded and your best homework will be given twice the weight. In addition, **there is an attendance requirement** as explained below under "Class Time."

# Collaboration

1. You can discuss the homework problems with classmates and on the discussion forum, and with the instructor and TA's during office hours, but you cannot discuss specific answers to a problem (neither selected nor excluded answers) among the given multiple-choice options with anyone at any time before the homework is due. This is a strict (honor code) rule.

**2.** There are no restrictions at all on discussing the contents of the lectures. After a homework or the final exam is due, you can discuss it freely anywhere with anyone.

**3.** No collaboration is allowed in the final exam. You can ask questions on the forum for clarification only.

# **Open Book and ChatGPT**

You can consult any books or notes, but there is an important exception: you cannot consult homework, final exam, and solutions from any source other than the class website (and then only according to the schedule on that website). The use of ChatGPT or other LLM's in this course is not allowed (there are numerous pages online with details about the course that may have been part of the training of these models).

# Class Time

The class will be held Tuesdays and Thursdays 2-4 PM in Beckman Institute Auditorium. In the first hour, a recorded lecture will be played. Attending the first hour is optional, as you can watch the video lecture ahead of time. The second hour is dedicated to a live discussion of the lecture with the instructor. You are <u>required</u> to attend the discussion part of at least 8 out of the 18 lectures during the term. Because of thanksgiving holiday, the lectures are on 10 Tuesdays and 8 Thursdays. Short presentations, called **New Take**, about some topics representing the state of the art in Machine Learning will be given live by the instructor during some of the discussion sessions.

## **Bonus Exercise**

There will be an optional bonus exercise about deep neural networks given shortly after midterm. Those who complete the exercise will have their attendance requirement reduced from 8 to 7 discussion sessions, and have their 2nd worst homework score augmented by 30 points (with the maximum being 100).