Machine Learning
(機器學習)

Course Introduction, 09/06/2023

Hsuan-Tien Lin (林軒田)

htlin@csie.ntu.edu.tw

Department of Computer Science
& Information Engineering

National Taiwan University
(國立台灣大學資訊工程系)

Four Reasons for NOT Taking the Course (1/4)

Complicated Contents

- from a Taiwanese student taking MIT ML class (translated):
  The professor started writing math equations as if he was using some writing accelerator. After class I always felt feeble. The worst part is: I needed to understand the contents as soon as I can. Otherwise I cannot finish the homework and cannot follow up in the next class.

- NTU ML class: designed to be **as good as** the best classes in the world

- similar things will happen to you

  If you are not willing to be so **miserable**, ...

Strict Instructor

- Will you give me a second chance if I copy homework from other people? **NO**.
- Could you let me pass because I will be kicked out by the $1/2$ rule? **NO**.
- Will you change my score from *F* to *C*? **NO**.
- How many will pass? **Any**, if necessary.

If you do not like a **strict** instructor, ...

Huge Loads

- from a student taking ML 2010 (posted on BBS):

lxxxxxx9 (translated): only 1 problem requires running 100 test experiments? (100*10min = 16hr); I double-checked multiple times—the one who designed this homework is **heartless**

- our class: **two to six** times harder than a normal one in NTU
- planning **six** homework sets (and final project)!
- homework due within 2 weeks in general
- will have homework 0 **in a few days**
  - already hard
  - very few points to remind you prerequisites

    If you do not want to spend **so much time on homework**, ...

Online-learnable

- invited by NTU as two of the **Massive Online Open Courses** on NTU-Coursera: **Machine Learning Foundations** and **Machine Learning Techniques**, which cover 87% of what we'd teach
- will generally ask you to watch those videos **at home before class** starting W2, and use in-class time for supplementary contents, discussion, etc.
- much **easier** to just learn online at home instead of **coming to class for discussions and suffering from homework loads**

If you want to **learn in your own pace**, ...

from a student in ML2013 (final feedback, translated): This is a class that makes you experience all kinds of devastation and desperation.

May the Brave Ones Stay

Basic Information

- mode: physical, with recordings if everything goes smoothly
- instructor: Hsuan-Tien Lin (htlin@csie.ntu.edu.tw)
- office hour: after class, or by appointment
- public course infopage: https://www.csie.ntu.edu.tw/~htlin/course/ml23fall/
- semi-private course platform: NTU COOL, for announcements, recorded videos, forum, etc.
- class chats: Discord (will announce on NTU COOL for invitation URL)
- homework grading (registered students only): Gradescope (will announce on NTU COOL for entry code)

> **be sure to check emails**
> **from NTU COOL frequently**

History of the Course

- 2008–2012: 3 credits, English-teaching
- 2013–2014: 3 credits, Mandarin-teaching (with MOOC)
- 2015: 4 credits, Mandarin-teaching (with MOOC)
- 2016–2020: mostly 2+2 credits (MLF+MLT), Mandarin-teaching (with MOOC)
- 2021: 3 credits, English-teaching
- 2023 Spring: 3 credits, English-teaching
- 2023 Fall: 3 credits, Mandarin-teaching (with MOOC **offline** and **physical discussions stimulated by chatGPT**)

> physical discussions stimulated by chatGPT:
> - many students prefer watching recorded lectures **offline** anyway
> - encourage **speculation** instead of **lecturing** during precious class time

Enrollment

- **Per school policy, students who have taken other versions is generally NOT allowed to take this class again!**
  (ML Foundations, ML Techniques, and my 4-credit ML)
    - we won't check, but **the school may**
    - multiple taking may not appear nice on your transcript anyway
- **limited number of students** (28 * # TAs = 308)
    - priority 0: CSIE/GINM/DS/Geo students (before 09/09)
    - priority 1: EECS students (before 09/11)
    - priority 2: NTU students (before 09/13)
    - priority 3: other students (before 09/15)
- auditing: welcomed in general if no interference to official members (check with TAs for accessing NTU COOL if needed)

> **Leave as soon as possible! Give your classmates a chance to be miserable.**

Our 11 Teaching Assistants
html_ta@csie.ntu.edu.tw

- Chia-Wei CHANG (Undergraduate in CSIE Department)
- Yu-Cheng CHENG (M.S. Student in CSIE Department)
- Shuo-Chen HO (Undergraduate in CSIE Department)
- Cai-Yi HU (M.S. Student in CSIE Department)
- Yu-Shiang HUANG (Ph.D. Student in Data Science Program)
- Ren-Wei (Willy) LIANG (Undergraduate in CSIE Department)
- Jeng-Yue (Buffett) LIU (Undergraduate in Geography Department)
- Poy LU (Ph.D. student in Graduate Institute of Networking and Multimedia)
- Odo To (Undergraduate in CSIE Department)
- Cheng-Chi (Casper) WANG (Undergraduate in CSIE Department)
- Hsuan-Fu WANG (M.S. Student in Graduate Institute of Networking and Multimedia)

THE Book

Learning from Data: A Short Course
Y. Abu-Mostafa (Caltech), M. Magdon-Ismail (RPI), H.-T. Lin (NTU)

- idea initiated during 2008
- 5 chapters, closely needed for first part of the class
- other e-Chapters to be used in the second part of the class
- teaching with the book and suggested reading within the book

Getting the Book to Read

- NTU Library: reserved copy in shared course material area
- R536: some shared copies to be read in the room (if open)
- Chuan-Hwa Book Company: imported some copies of the book
  —e.g. https://www.books.com.tw/products/0010565319
- Amazon/Kindle: main selling channel in the US
  — http://www.amazon.com/gp/product/1600490069

> **If the book is not affordable to you but you really want to read it: email me (htlin@ csie.ntu.edu.tw) and I'll see how I can help.**

THE Principle

Taking any unfair advantages over other class members is not allowed. It is everyone's responsibility to maximize the level of fairness.

NO CHEATING

NO LYING

NO PLAGIARISM

NO PIRATING of THE BOOK

very serious consequences

Grade

- no midterm, no final exam
- main reference: homework sets + final project
- raw score goes through some order-preserving normalization
  - raw score 80 with term rank A: possible
  - raw score 60 with term rank F: possible
  - raw scores 80, 60 with term scores B, B: possible, but unlikely
  - raw scores 80, 60 with term scores F, B: **impossible**

Collaboration and Open-Book

- homework discussions: encouraged
- but fairness?
  write the final solutions alone and understand them fully
- references (books, notes, Internet **[chatGPT]**):
  can be consulted, but **not copied from**
- no need to lend/borrow solutions

> to maximize fairness
> (everyone's responsibility),
> **lending/borrowing not allowed**

Collaboration and Open-Book

to maximize fairness (everyone's responsibility),
lending/borrowing not allowed

Deal? If your classmate wants to borrow homework from you, what do you say?

Homework

- 12+1 problems (**fewer**, yeah!!) per homework, TA-graded
- upload solutions and code to gradescope
- penalty for late parts:
  90% of value for 12-hour late, 80% one-day late, ...

> will grant each person four penalty-free late
> half-days (**gold medals**)

Programming Assignments

- about a third or half of the problems
- any programming language, any platforms
- **no sophisticated packages**

> students' responsibility: **ask TA in advance for what can/cannot be used**

Experimental Items for HTML2023Fall

- at-home lecture watching
- physical discussions stimulated by chatGPT
- returning to hand-written homework (with fewer problems) and full TA grading

> **Feel free to give us feedback (immediately or afterwards) to help improve the course!**

Questions?