# Machine Learning
# (機器學習)

## Course Introduction, 09/02/2024

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, ...

# Four Reasons for NOT Taking the Course (2/4)

## 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, ...

# Four Reasons for NOT Taking the Course (3/4)

## 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 seven homework sets (and final project)!
- homework due within 2 weeks in general
- will have homework 0 this week
  - already hard
  - very few points to remind you prerequisites

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

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

## 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 60% of what we'd teach, in Mandarin
- much easier to just learn online at home—you can simply take the recorded version instead if you know Mandarin
- actually, there is sister course at Caltech in English

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 live Youtube screencast + Slido (#HTML2024FALL) questions 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/ml24fall/`
- semi-private course platform: NTU COOL, for announcements, recorded videos, etc.
- class chats: Discord (see NTU COOL for invitation URL)
- homework grading (registered students only): Gradescope (linked from NTU COOL)

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

aims of doing English-teaching again:

- be more language-friendly to international students
- keep the course evolving

# Three New Experiments this Semester (1/3)

## cross-university teaching

- issue: Taiwan needs more AI talents
- accepted invitation from TAICA (Taiwan AI College Alliance) to offer this ML course to 25 universities
  - thanks to MOE for more TAs
  - YouTube screencast to break the time/space boundary
  - some online TA hours (to be announced)

cross-university teaching: attempt to further increase the impact of this course in this new AI age

# Three New Experiments this Semester (2/3)

## better regrading protocol

- issue: student-TA misunderstanding when regrading over Gradescope
- 2 back-and-forths
  $\Longrightarrow$ appealing to VP of Assignment (TA Chris Chien)
  ($\Longrightarrow$ HT's attention if needed)
  - more back-and-forths are less useful
  - third-party involvement to resolve disputes effectively

better regrade protocol: make regrading about increasing mutual understanding, rather than adversarial disputes

# Three New Experiments this Semester (3/3)

## concise in-class question-answering

- issue: overly-lengthy QA sessions in class, possibly worsen with chatGPT-assisted answering and our cross-university population
- answer 3-5 physical/top questions per session online only; other questions will be answered offline
- still welcomed to reach me after class for physical discussions

concise in-class QA: balance class pace and our exciting discussions

# 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
- almost unlimited number of NTU students (type-3 with quota of 450)
- limited quota for Taiwan AI College Alliance students from other universities
- 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 18 Teaching Assistants

html_ta@csie.ntu.edu.tw

- Yun-Ye Cai
- Chia-Wei Chang
- Shu-Han Chang
- Chia-Le Chen
- Chu-Hsin Chen
- Chien-Yi (Chris) Chien
- Shih-Hsuan Chou
- Mai Tan Ha
- Bo-Kai Huang
- Chien-Jui Huang
- Hsiao-Chieh Kao
- Yu-Wei Kuan
- Hsun-Yu (Yoyo) Lee
- I-Pei Lee
- Ren-Wei Liang
- Zhi-Bao Lu
- Wei-Po Wang
- Chun-Hao Yang

# 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
- for non-NTU students, (tentatively) the rank will be converted by NTU metric back to a score between $[0, 100]$, and then your university will convert it back to the rank

# 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/sell/buy/trade solutions

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

# Homework

- 12+1 problems per homework, human-graded
- upload solutions and code to gradescope with problem tagging to facilitate TAs' grading
- 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

# Languages

- teaching: English
- Slido questions: English/Traditional Chinese/Simplified Chinese; will be translated by instructor before answering
- homework writing: English-only
- TA hour/instructor hour: English/Mandarin; all TA hours English-welcomed, while having a few English-prioritized TA hours
- TA Email: English-only
- Discord: separate English-only channels and Mandarin+English channels

don't be afraid of English

# Questions?