2022 機器學習
課程規定

李宏毅
Hung-yi Lee
About this course

• Time slot: 2:20 p.m. – 6:20 p.m., Friday
• Classroom: 博理 112
  • Live streaming during the lecture time
  • All lectures will be recorded
• You can complete this course online.
  • submit homework online, no exam
• Prerequisite
  • Math: Calculus (微積分), Linear algebra (線性代數) and Probability (機率)
  • Programming: You can read and write python code.
About this course

• Focus on **deep learning**
  • Can be your first machine learning (ML) course.
  • Little overlap with Hsuan-Tien Lin’s (林軒田) *Machine Learning Foundations* and *Machine Learning Techniques*.

---

Diagram:

- Hsuan-Tien’s ML
- This course
  - Computer Vision
  - Speech
  - Natural Language
- your first ML course
About this course

• Focus on **deep learning**
  • Can be your first machine learning (ML) course.
  • Little overlap with Hsuan-Tien Lin’s (林軒田) *Machine Learning Foundations* and *Machine Learning Techniques*.

• Covering **broad aspects**
  • Try to cover most important technology and concepts you need to know (buffet style!)
  • Not delve into most topics. This is your first ML course, not the last one.

• Covering **the latest technology**
• **Application oriented**
Applications

**COVID-19**

- HW1

**Computer Vision**
- CNN
- HW 3, 8, 9, 10, 11
- attack, adaptation, compression, explanation, anomaly detection
- Image generation
- HW 6

**Natural Language Processing**
- HW 5
- HW 7

**Speech Processing**
- HW 2
- HW 4

**RL**
- HW 12
Webpage

• All the recording and assignments will be available on the course webpage.

• Course webpage:
  https://speech.ee.ntu.edu.tw/~hylee/ml/2022-spring.php
Assignment
Assignment

• Most assignments include report, leaderboard, and code submission.
  • Report: answer some questions
  • Leaderboard (排行榜): Kaggle or JudgeBoi (our in-house Kaggle 😊)
    • Simple, medium, strong, boss baselines
  • Submit the related codes of each assignment via NTU COOL.

• All assignments can be done by Google Colab. You can pass this course without preparing hardware or install anything.

• But usually more computing resources lead to better performance.
Grading Criterion

- There are **15** assignments.
- Each has **10 points**, only count the **10** assignments with the highest points.
- You don’t need to do all the assignments. Choose the ones you are interested in.
- You are encouraged to complete all **15** assignments!

You decide how much you want to learn. **It’s buffet style.**

分數與等第換算按照學校建議
Disclaimers

• This course will NOT teach Python.
• This course will NOT teach any Python package, except PyTorch.
• Only focus on ML. TAs do not have to answer questions not related to ML or PyTorch.
• All TAs’ sample codes can be run on Colab. If you use your own device, TAs have no obligation to solve all problems.
• TAs have no obligation to help you pass the baselines.
• This course will NOT provide computing resources.

• When it comes to network training, your efforts are not always proportional to your performance.
• Network training can take a long time.
Lecture Schedule
課程網站

[課程網站](https://speech.ee.ntu.edu.tw/~hylee/ml/2022-spring.php)

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topic</th>
<th>Preparation - zh</th>
<th>Preparation - en</th>
<th>Class Videos</th>
<th>Class Notes</th>
<th>Extra Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>第三週</td>
<td>3/04</td>
<td>Image as input</td>
<td>Video</td>
<td>Video</td>
<td>#</td>
<td>CNN: <a href="#">ppt/pdf</a></td>
<td>Spatial Transformer Layer: <a href="#">Video</a></td>
</tr>
</tbody>
</table>

每一個主題都有一個對應的作業
Lecture Schedule

• Watch assigned videos before the lecture
• During lecture
  • Teach something new (usually 1 hour) or invited speakers
    • Not directly related to assignments
  • Assignment announcement by TA
  • We will usually finish the lectures before 6:20 p.m.
• You can complete this course online.
Kaggle
Kaggle (JudgeBoi is similar)
https://www.kaggle.com/

- Some assignments are in-class competition on Kaggle.
- Register a Kaggle account by yourself.
<table>
<thead>
<tr>
<th>#</th>
<th>Team Name</th>
<th>Notebook</th>
<th>Team Members</th>
<th>Score</th>
<th>Entries</th>
<th>Last</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>b06902021_rm -f trained_model</td>
<td></td>
<td></td>
<td>0.77550</td>
<td>38</td>
<td>9mo</td>
</tr>
<tr>
<td>2</td>
<td>b05901176__/6_/_/</td>
<td></td>
<td></td>
<td>0.77400</td>
<td>28</td>
<td>9mo</td>
</tr>
<tr>
<td>3</td>
<td>b05901063_QQ</td>
<td></td>
<td></td>
<td>0.77380</td>
<td>23</td>
<td>9mo</td>
</tr>
<tr>
<td>4</td>
<td>r07522839_劉承岳</td>
<td></td>
<td></td>
<td>0.77130</td>
<td>11</td>
<td>9mo</td>
</tr>
<tr>
<td>5</td>
<td>b06902030_5/14資訊之夜</td>
<td></td>
<td></td>
<td>0.77020</td>
<td>30</td>
<td>9mo</td>
</tr>
<tr>
<td>6</td>
<td>b04901147_系吃隊長</td>
<td></td>
<td></td>
<td>0.76920</td>
<td>24</td>
<td>9mo</td>
</tr>
<tr>
<td>7</td>
<td>r07943150_ML靠賽 輕鬆自在</td>
<td></td>
<td></td>
<td>0.76830</td>
<td>35</td>
<td>9mo</td>
</tr>
<tr>
<td>8</td>
<td>r07943156_慈母守中線遊子遊野...</td>
<td></td>
<td></td>
<td>0.76770</td>
<td>19</td>
<td>9mo</td>
</tr>
</tbody>
</table>

Your submission scored 0.76920, which is not an improvement of your best score. Keep trying!
Kaggle

• The display name should be

   `<STUDENT ID>_ <ANY THING>`

   b93901106  truly any thing 😊

• Example

   O  b93901106_pui pui pui pui pui pui pui pui
   O  b93901106_
   X  b93901106 puipui

We will not find your submission if your format is wrong!
Public score: You can see it right after the submission.

Private score: You can only see the score after the assignment deadline.

<table>
<thead>
<tr>
<th>#</th>
<th>Team Name</th>
<th>Notebook</th>
<th>Team Members</th>
<th>Score</th>
<th>Entries</th>
<th>Last</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>b06902021_rm -f trained_model</td>
<td></td>
<td></td>
<td>0.77550</td>
<td>38</td>
<td>9mo</td>
</tr>
<tr>
<td>2</td>
<td>b05901176_(•ॹ•ॹ)(/)</td>
<td></td>
<td></td>
<td>0.77400</td>
<td>28</td>
<td>9mo</td>
</tr>
<tr>
<td>3</td>
<td>b05901063_QQ</td>
<td></td>
<td></td>
<td>0.77380</td>
<td>23</td>
<td>9mo</td>
</tr>
<tr>
<td>4</td>
<td>r07522839_劉承岳</td>
<td></td>
<td></td>
<td>0.77130</td>
<td>11</td>
<td>9mo</td>
</tr>
<tr>
<td>5</td>
<td>b06902030_5/14資訊之夜</td>
<td></td>
<td></td>
<td>0.77020</td>
<td>30</td>
<td>9mo</td>
</tr>
<tr>
<td>6</td>
<td>b04901147_系吃隊長</td>
<td></td>
<td></td>
<td>0.76920</td>
<td>24</td>
<td>9mo</td>
</tr>
</tbody>
</table>

Your Best Entry:
Your submission scored 0.76920, which is not an improvement of your best score. Keep trying!
Kaggle – Pokémon & Digimon

**Testing Data**

<table>
<thead>
<tr>
<th>Ground truth</th>
<th>Given in the assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>poki</td>
<td>poki</td>
</tr>
<tr>
<td>poki</td>
<td>digi</td>
</tr>
<tr>
<td>poki</td>
<td>poki</td>
</tr>
<tr>
<td>digi</td>
<td>digi</td>
</tr>
<tr>
<td>digi</td>
<td>digi</td>
</tr>
</tbody>
</table>

Given on Kaggle (unseen)
Kaggle – Pokémon & Digimon

Ground truth
- pokí
- pokí
- digi

Model Prediction
- pokí
- digi
- digi

public

private

Acc = 2/3
What you can see immediately

Acc = 1/3
After the submission deadline
• You need to select two results for evaluating on the private set before the assignment deadline.

• You only have limited submission times per day.
Rules
Rules – Common Sense

• Don’t plagiarize others’ code and don’t submit others’ reports or results.
  • “Other” means all creatures in the universe
  • Using the available public toolkits is allowed.
  • If some of your codes are from others’ repositories, please mention them in your code.
  • If you discuss your assignments with some classmates/friends, mention them in your code.
  • TAs and the lecturer decide plagiarism or not.
Rules – Common Sense

- Protect your efforts! Don’t let others see your codes, don’t give others your results.
  - Lending your codes to others or allowing others to copy your work will be considered as collusion, thus receiving the same punishment as the plagiarist.
Rules – For Kaggle and JudgeBoi

• There is a limited number of submissions to all the leaderboards (Kaggle and JudgeBoi).
  • Don’t try to have multiple accounts. (It also violates the rules of Kaggle.)
  • Don’t borrow account from others and don’t give you account to others.
  • Don’t submit your results to leaderboards of previous courses.
  • Don’t use any approach to increase the submission numbers
Rules – For Kaggle and JudgeBoi

• The results submitted to the leaderboards should **only come from machines**.
  • Don’t label the testing data by humans (or any other approaches)!

• Only use the data provided in each assignment!
Rules - Codes

• You need to submit codes for each assignment via NTU COOL.

• Your codes need to be able to generate the results you submit to the leaderboard.
  • If not, it would be considered cheating and get punishment.
  • TAs may not run all the codes, but TAs will check some of them.
  • TAs and the lecturer decide cheating or not.
Punishment

• The **first time** you violate the rules.
  • The final score of this semester times 0.9, and you receive zero score for the assignment you violate the rules.

• The **second time** you violate the rules.
  • Fail the course.
加簽
加簽

• 加簽電資學院(含輔系*、資料科學學程、智慧醫療學程)的學生
  • 請填寫 google 表單 (如果沒有要加簽就不要填、也不要幫其他人填)
  • 等一下大助教會公告表單連結
  • 表單填寫期限到下週三(2/23)午夜，逾時不候
  • 之後會透過 NTU COOL 發授權碼

*輔系指事前核可的輔系
加簽

• 非電資學院的學生，依據作業一的 leaderboard 排名加簽 (等一下助教會公告作業一)
  • 也請填寫 google 表單
  • 根據 private leaderboard 排名取前 30 名非電資學院學生獲得加簽資格
    • 不看 public leaderboard
    • 只看排名不看分數
  • 不遞補
  • 之後會透過 NTU COOL 發授權碼
• 無論是否為電資學院的學生，完成作業一都可以計入期末總分
旁聽

• 本課程歡迎旁聽
• 課程內容和作業內容都已經完整公開在課程網頁上，有沒有正式修課對於學習影響不大
• 旁聽生請寄信給助教，可以加入 NTU COOL
• 旁聽生可以上傳結果到 Kaggle (但無法上傳到 JudgeBoi)
• 助教不批改旁聽生的報告
Questions?
Questions

• **Option 1**: Ask at TA hour

• **Option 2**: Post your questions on NTU COOL
  - Your questions are also other’s questions.

• **Option 3**: Mail to the following address
  - E-mail: mlta-2022-spring@googlegroups.com
  - E-mail title includes “[hwX]” (e.g. [hw3])

• Don’t direct message to TAs. The TAs will only answer the questions by the above alternatives.
TA email: mlta-2022-spring@googlegroups.com

TA head  陳子晴