CS5284 : Graph Machine Learning

Administrative (Week 2)

Semester 1 2025/26

Xavier Bresson

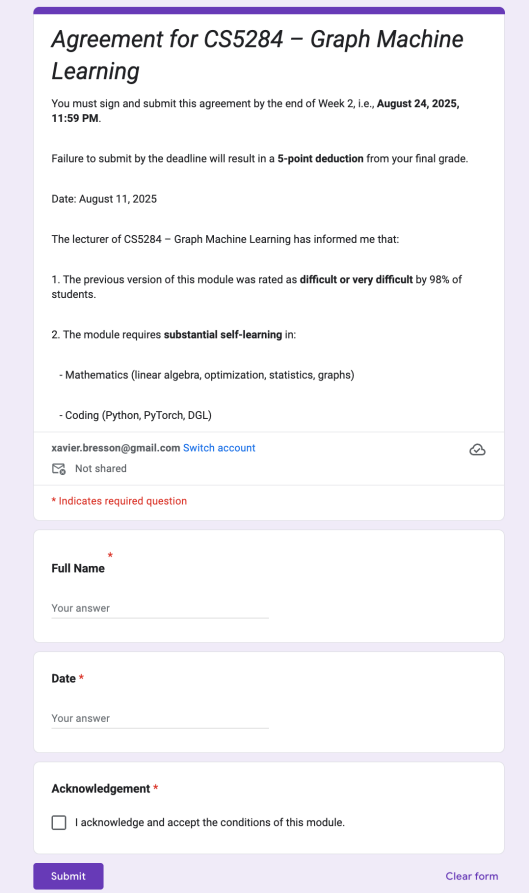
https://x.com/xbresson

Department of Computer Science
National University of Singapore (NUS)

Admin

Agreement

- Please, fill out the agreement : https://forms.gle/tt1bxQA4f2CxGEYW9
- Deadline : End of Week 2, August 24, 2025, 11:59pm.
- Penalty : 5-point deduction from the final grade.



Agreement for CS5284 – Graph Machine Learning

You must sign and submit this agreement by the end of Week 2, i.e., **August 24, 2025, 11:59 PM**.

Failure to submit by the deadline will result in a **5-point deduction** from your final grade.

Date: August 11, 2025

The lecturer of CS5284 – Graph Machine Learning has informed me that:

1. The previous version of this module was rated as **difficult or very difficult** by 98% of students.

2. The module requires **substantial self-learning** in:

- Mathematics (linear algebra, optimization, statistics, graphs)

- Coding (Python, PyTorch, DGL)

xavier.bresson@gmail.com Switch account

Not shared

* Indicates required question

Full Name *

Your answer

Date *

Your answer

Acknowledgement *

I acknowledge and accept the conditions of this module.

Submit Clear form

Student questions

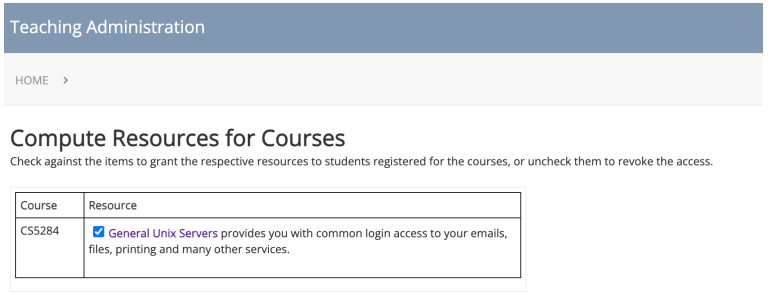
Student question

- Following last week's discussion, here are a few books for the module (and beyond) :
  - Introduction to Linear Algebra, Gilbert Strang : https://www.dropbox.com/scl/fi/i1ofqcfgmggbgnq43ajpj/linear_algebra_4th_strang.pdf?rlkey=lzjf51q6g7y63s6g512okkr3w&dl=0
  - Spectral Graph Theory, Fan Chung : https://www.dropbox.com/scl/fi/dcdd0cub3mpa9sv76gtne/spectral_graph_theory_fan_chung.pdf?rlkey=zdcoumx2oz7rzb6iv5v9079uz&dl=0
  - Deep Learning, Ian Goodfellow and Yoshua Bengio and Aaron Courville : https://www.dropbox.com/scl/fi/v71bhr25nj4nrn5ss4b57/deep_learning_goodfellow_bengio_courville.pdf?rlkey=k8cmlplucziba5pfvl93kqv87&dl=0
  - The Elements of Statistical Learning, Hastie, Tibshirani, Friedman : https://www.dropbox.com/scl/fi/u66b0zy8xhcpjhtejinwz/the_elements_of_statistical_learning_hastie_tibshirani_friedman.pdf?rlkey=b9b3s9tsdoh1fnn80e3odma1e&dl=0
- This slide will auto-destruct in a week ☺

Student question

Will any GPU resource be provided for the course project?

- GPU for the project
  - Free GPU with Google Collab
  - Google Cloud platform offers 150 hrs of free GPU
  - School GPU (but I need to figure out more)
- Will explain the project on W6.

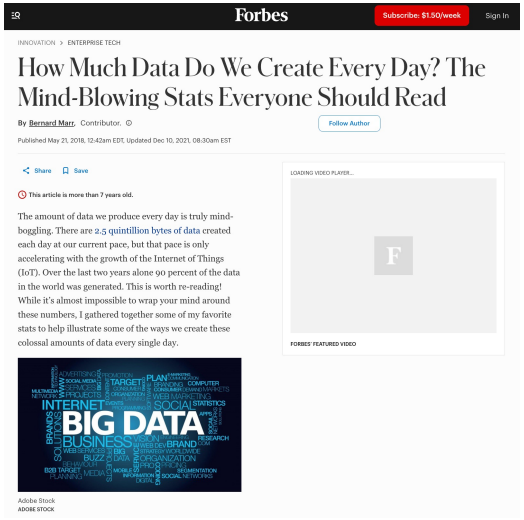


Student question

In the lecture 1 slide, it was mentioned that "Google's Law of Data : Volume of data double every 1.5 years." and Eric Schmidt : "Every 2 days we create as much information as we did up to 2003." But for the 'volume of data' that was mentioned here, did it count the duplicate data, or the data that essentially means the same thing? Would you be willing to share something on how this 'data volume' and 'amount of information' were calculated?

- It is a rule of thumb (not rigorously proved) to capture the idea that the 3rd industrial revolution (the digital one) has produced and will increasingly generate more and more digital data.
- One source :

  https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read



Student question

In the lecture 1 slide, it was mentioned that "Google's Law of Data : Volume of data double every 1.5 years." and Eric Schmidt : "Every 2 days we create as much information as we did up to 2003." But for the 'volume of data' that was mentioned here, did it count the duplicate data, or the data that essentially means the same thing? Would you be willing to share something on how this 'data volume' and 'amount of information' were calculated?

It was mentioned in the 'supervised learning' slide. Does the law apply to generally the data that exist on the internet (for supervised learning), or does it apply to what could be scrapped from the internet for LLM training etc?

- Great question!
- A new trend has emerged since the introduction of LLMs and GenAI.
- Before 2022, duplicates may have been a concern, but since then, generated data has flooded the Internet!
- In April 2025, an analysis of 900,000 English-language web pages showed that 74.2% of newly created pages included AI-generated content.

https://www.stanventures.com/news/new-web-pages-now-contain-ai-generated-content-2727

Questions?