

COM6012: **Scalable** Machine Learning

Spring 2025

University of Sheffield

<https://github.com/com6012/ScalableML>



Created by DALL-E

Check-in code:
XX-XX-XX

It seems like a good idea, but is it scalable?

https://s3.amazonaws.com/lowres.cartoonstock.com/animals-scalable-product-mice-cats-slingshot-mdbn347_low.jpg



CS423878

Three Instructors + Guest (X. Liu)



Shuo Zhou
Module lead



Tahsin Khan

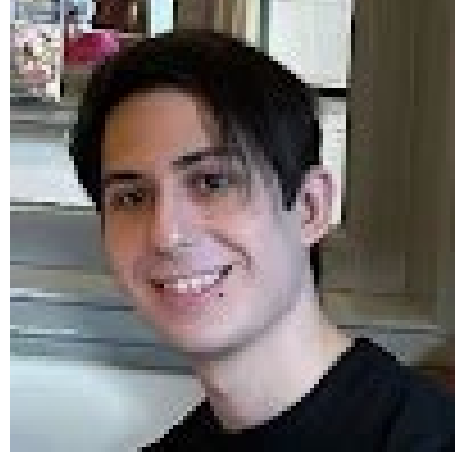


Haiping Lu

Five Demonstrators (TAs)



Areeb Sherwani (Head)



Christopher J
Noroozi



Charlie Grimshaw



Xiaolei Xu



Xiaozhou Tan

Lectures and Labs

Lecture @ Diamond LT5

Wednesday
12:00–13:00

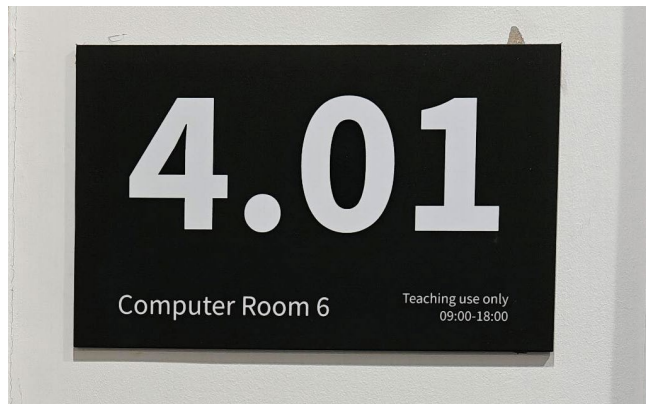
Thursday
11:00–13:00

Lab @ Diamond Computer
Room 6 (4.01)

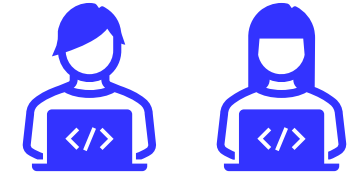


Labs

- Bring your laptop if needed
- Use HPC clusters
- Finish required tasks of the labs in this slot!



Contents: **Very Hands-on**



Week	Date	Topic	Instructor
1	12 Feb	Introduction to Spark and HPC	Shuo
2	19 Feb	RDD, DataFrame, ML pipeline, & parallelisation	Shuo
3	26 Feb	Scalable logistic regression	Shuo
4	05 Mar	Scalable generalised linear models	Shuo
5	12 Mar	Scalable decision trees	Tahsin
6	19 Mar	Scalable neural networks	Tahsin
7	26 Mar	Scalable K-means clustering	Tahsin
8	2 Apr	Scalable matrix factorisation for collaborative filtering (RecSys) and PCA for dimensionality reduction	Haiping
9	30 Apr	Apache spark in the cloud (not assessed)	Xianyuan
10	7 May	Reproducible and reusable AI (not assessed)	Xianyuan

Assessment

- Lab exercises: 0% (self assessment)
 - Finish lab exercises by the following Tuesdays
 - Solutions to release on the following Wednesdays
- Assignment: 40%
 - HPC driving licence (1%) – complete before first lab on 13th Feb (50% penalty for late submission)
 - Formal assignment (39%) – **avoid using unfair means**
 - Progressive release, to be handed out by 27th March
 - Deadline: 13:00 on Thursday, **8th May** (end of lab)
 - Solution release: 23rd May
 - Marking and feedback deadline: 30th May
- Exam: 60%
 - To be scheduled in exam period (26th May to 14th June)
 - Formal exam on Blackboard: 2 hours (with samples available)

Additional Sessions and Support

- Additional sessions

- Instructor office hour: Wednesday 5:00-6:00 pm @G25 Regent Court
- Online Discord help sessions: Thursday 2pm-3pm from week 2

- Blackboard discussion board

- One general forum: general question/feedback
- Three forums: Week 1-4, Week 5-7, and Week 8
 - Get help on lecture/lab contents
 - To ask for **clarification** on assignment **questions** (i.e. the tasks to do)
 - **NOT to ask how to solve the problems.**
 - **NOT to ask for the correctness of a specific solution, or share a possible solution.**

- Direct email to instructors: **personal/private issues only**