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

# COM6012: Scalable Machine Learning - Spring 2024

https://github.com/com6012/ScalableML

University of Sheffield

# Three Instructors + Guest (X. Liu)

Shuo Zhou

Module leader

Robert Loftin

Tahsin Khan

# Four Demonstrators (TAs)

Areeb Sherwani (Head)

Pawel Pukowski

Sina Tabakhi

Xiaolei Xu

# Lectures and Labs

Lecture: Monday 1pm-2pm @ Pam Liversidge Building LT02

**Monday, 1:00 PM –2:00 PM**

**Friday, 11:00 AM –1:00 PM**

Lab: Friday 11am-1pm @ Diamond 207

# Labs

- **Bring your own laptop**
  - Laptop loans are available on level 4 of the Diamond; please use the lockers located in front of room 4.17's entrance


- Finish required tasks of the labs in this slot!

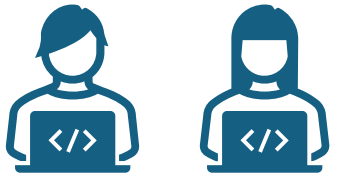
# Other Interactions

- Additional sessions
  - Instructor office hour: Mon 4:30-5:30 pm @G25 Regent Court
  - Online Discord help sessions: Wed 2pm-3pm from week 2
- Blackboard discussion board: post your questions to get answered by the course instructors
  - One general forum: general question/feedback
  - Three forums: Shuo, Robert, and Tahsin
    - Get help on lecture/lab contents
    - Ask respective assignment questions
  - Assignment questions:
    - To ask for clarification on assignment questions (i.e. the tasks to do)
    - NOT to ask how to solve the problems, the correctness of a specific solution, or share a possible solution. It is an assessment.
- Direct email to instructors: **personal/private issues only**

# Assessment

- Lab exercises: 0% (self assessment)
  - Finish lab exercises by the following Wednesdays
  - Solutions to release on the following Thursdays
- Assignment: 50%
  - Progressive release → start EARLY: complete by 22 March?
  - Deadline: 13:00 on Fri, 3rd May (end of lab)
  - Solution release: 13th May
  - Marking and feedback deadline: 27th May
- Final exam: 50%
  - To be scheduled in exam period (20th May – 8th June)
  - Formal exam on Blackboard: 2 hours (with samples available)

# Contents: Very Hands-on



| Week | Date | Topic | Instructor |
|---|---|---|---|
| 1 | 05 Feb | Introduction to Spark and HPC | Shuo |
| 2 | 12 Feb | RDD, DataFrame, ML pipeline, & parallelization | Shuo |
| 3 | 19 Feb | Scalable logistic regression | Shuo |
| 4 | 26 Feb | Scalable generalized linear models | Robert |
| 5 | 04 Mar | Scalable decision trees | Tahsin |
| 6 | 11 Mar | Scalable neural networks | Tahsin |
| 7 | 18 Mar | Scalable matrix fact. for collaborative filtering (RecSys) | Robert |
| 8 | 15 Apr | Scalable K-means clustering | Robert |
| 9 | 22 Apr | Scalable PCA for dimensionality reduction | Robert |
| 10 | 29 Apr | Apache spark in the cloud (guest lecture, not assessed) | Xianyuan |