It seems like a good idea, but is it scalable?
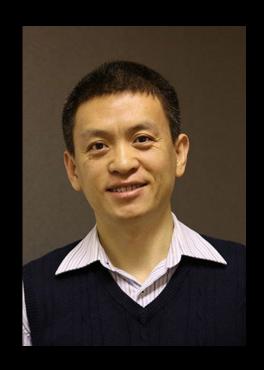
https://s3.amazonaws.com/lowres.cartoonstock.com/animals-scalable-product-mice-cats-slingshot-mdbn347_low.jpg

COM6012: Scalable Machine Learning - Spring 2023

https://github.com/haipinglu/ScalableML (Since 2019)

University of Sheffield

Three Instructors + Guest (M. Smith)

Haiping Lu
Module leader

Shuo Zhou

Tahsin Khan
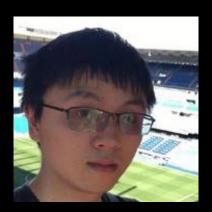
Four Demonstrators (TAs)

Areeb Sherwani (Head)

Pawel Pukowski

Sina Tabakhi

Peizhen Bai

Lectures and Labs

- Lecture: Monday 1pm-2pm @ Diamond LT2
- Lab: Friday 11am-1pm @ Diamond Computer Room 1
  - Finish required part of the labs in this slot!

Other Interactions

- Additional sessions
  - Instructor office hour: Mon 4:30-5:30pm @G25 Regent Court
  - Online Discord help sessions: Wed 2pm-4pm from week 2
- Blackboard discussion board: post your questions to get answered by the course instructors
  - One general forum: general question/feedback
  - Three forums: Haiping, Shuo, 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 31 March
  - Deadline: 13:00 on Fri, 5th May (end of lab)
  - Solution release: 12th May
  - Marking and feedback deadline: 26th May

- Final exam: 50%
  - To be scheduled in exam period (22 May – 9 June)
  - Formal exam on Blackboard: 2 hours (with samples available)

Contents: Very Hands-on

Week	Date	Topic	Instructor
1	06 Feb	Introduction to Spark and HPC	Haiping
2	13 Feb	RDD, DataFrame, ML pipeline, & parallelization	Haiping
3	20 Feb	Scalable logistic regression	Shuo
4	27 Feb	Scalable generalized linear models	Shuo
5	06 Mar	Scalable matrix fact. for collaborative filtering (RecSys)	Haiping
6	13 Mar	Scalable K-means clustering	Haiping
7	20 Mar	Scalable PCA for dimensionality reduction	Haiping
8	27 Mar	Scalable decision trees	Tahsin
9	*24 Apr*	*Apache spark in the cloud (guest lecture, not assessed)*	*Mike*
12	*15 May*	*Scalable neural networks (not assessed)*	*Tahsin*