Introduction to Data Science & Machine Learning

Reusing this material

archer2

epcc

THE UNIVERSITY OF EDINBURGH

Course material

- Material primarily developed by Adam Carter, Ioanna Lampaki, and Michael Bareford at EPCC
  - Contributions by others as well

- UK National Supercomputer Service
  - managed by UKRI/EPSRC
  - Housed, operated and supported by EPCC
  - hardware supplied by Cray
- Training provided by the ARCHER2 Computational Science and Engineering (CSE) support team
  - 60 days per year at various locations
  - free to all academics

archer2

epcc

THE UNIVERSITY OF EDINBURGH

Located at EPCC's Advanced Computing Facility (ACF)

archer2

epcc

THE UNIVERSITY OF EDINBURGH

Key ARCHER2 Resources

- Upcoming courses
  - http://www.archer2.ac.uk/training/
- Material from past courses
  - https://www.archer2.ac.uk/training/materials/
- Virtual tutorials (online)
  - http://www.archer2.ac.uk/training/
- Documentation
  - http://www.archer2.ac.uk/documentation/

Who am I?

Adrian Jackson a.jackson@epcc.ed.ac.uk

- Teach at EPCC on HPC and program optimisation:
  - MSc, PhDs
  - PRACE Advanced Training Centre
  - ARCHER training programme
  - commercial training
  - ...
- Also do HPC research
  - Porting and optimizing community codes
  - new parallel programming models, accelerators, performance, memory hardware, ..

Other Resources

- Please fill in the feedback form!
  - you will be sent a link at the end of the course
- General enquiries about ARCHER2 go to the helpdesk
  - support@archer2.ac.uk
- EPCC runs one-year taught postgraduate masters courses
  - ***MSc in HPC*** and ***MSc in HPC with Data Science***
  - awarded by the University of Edinburgh since 2001
  - scholarships available
  - http://www.epcc.ed.ac.uk/msc/

Access during course

- Personal accounts for duration of course
  - will allow machine access for up to a month afterwards
- Accounts will be closed two weeks after access ends
  - all files etc. will be deleted
  - take copies of all your work beforehand!
- Course materials (slides, exercises etc) will continue to be available from ARCHER2 website
  - archived on ARCHER2 training pages for future reference

Code of Conduct

https://www.archer2.ac.uk/about/policies/code-of-conduct.html

- We expect all course trainers and attendees to:
  - Use welcoming and inclusive language
  - Be respectful of different viewpoints and experiences
  - Gracefully accept constructive criticism
  - Focus on what is best for the community
  - Show courtesy and respect towards other community members
- See web page for full details and incident reporting form

Timetable

- Tuesday 12th August
  - 10:00 Welcome & Introduction
  - 10:10 Lecture: Introduction to Data Science
  - 10:50 Practical: Intro to Jupyter Notebooks & Pandas
  - 11:30 Break
  - 11:45 Lecture: Introduction to Machine Learning
  - 12:15 Lecture: Linear Regression
  - 13:00 Lunch
  - 14:00 Practical: Linear Regression
  - 14:30 Lecture: kNN Classification
  - 15:00 Practical: kNN Classification
  - 15:15 Break
  - 15:30 Lecture: KMeans Clustering
  - 16:00 Close
- Wednesday 13th August
  - 10:00 Welcome and recap
  - 10:10 Practical: KMeans Clustering
  - 10:30 Lecture: Introduction to Neural Networks
  - 11:00 Practical: Simple Neural Network
  - 11:30 Break
  - 12:00 Lecture: More Neural Networks
  - 12:30 Practical: Simple Neural Network (Part 2)
  - 13:00 Lunch
  - 14:00 Practical: A Neural Network with PyTorch
  - 15:00 Wrap-Up and Questions
  - 16:00 Close

Course material

- https://github.com/EPCCed/archer2-intro-ds-and-ml

I hope you enjoy the course

- ... and *please ask questions*!