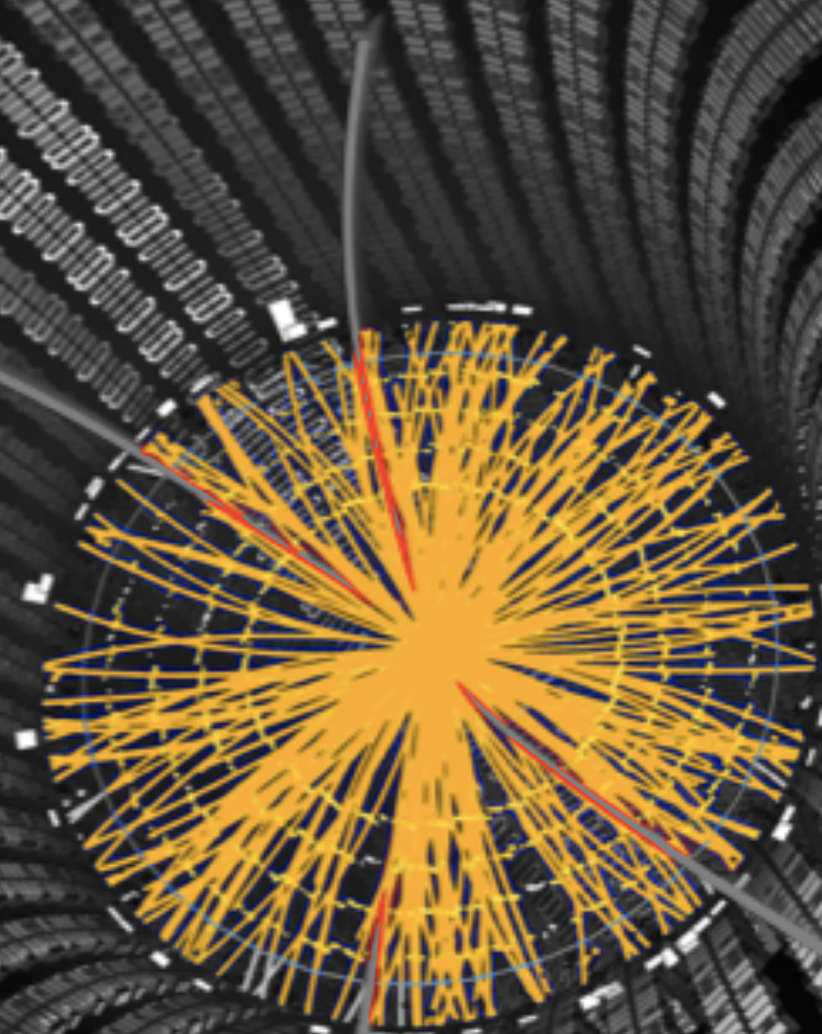
Bridging High-Energy Physics and Machine Learning communities

9 - 13 November 2015, CERN

DS@LHC



Local Organising Committee

- Xabier Cid (CERN)
- Gilles Louppe (CERN)
- Michelangelo Mangano (CERN)
- Maurizio Pierini (CERN)
- Jean-Roch Vlimant (Caltech)

Program Committee

- Kyle Cranmer (New York U)
- Cécile Germain (LRI)
- Vladimir Vava Gligorov (CERN)
- Gilles Louppe (CERN)
- Andrew Lowe (Wigner RCP)
- Maurizio Pierini (CERN)
- David Rousseau (LAL-Orsay)
- Maria Spiropulu (Caltech)
- Jean-Roch Vlimant (Caltech)
- Daniel Whiteson (UC Irvine)

International Advisory Committee

- Roger Barlow (Huddersfield U)
- Tommaso Dorigo (INFN-Padova)
- Ian Fisk (Simons Foundation)
- Maria Girone (CERN)
- Eilam Gross (Weizmann)
- Balázs Kégl (LAL-Orsay)
- Constantin Loizides (LBNL)
- Stuart Russell (UC Berkeley)
- Victoria Stodden (UI Urbana-Champaign)
- Max Welling (Amsterdam U)

sponsored by

LHC Physics Center at CERN: http://lpcc.web.cern.ch
Fermilab National Laboratory: http://fnal.gov
Moore-Sloan Data Science Environment: http://cds.nyu.edu/mooresloan

http://cern.ch/DataScienceLHC2015

HEP vs the outside world

- Traditionally, HEP community use advanced data analysis tools taken from other fields
  - Maximum Likelihood fits in the 80's
  - Neural Networks in the 90's
  - BDTs in the 00s
  - ….
- Traditionally, we also make our own tools
  - Minuit, PAW, ROOT, RooFit, TMVA, …
  - Only rarely our tools penetrate the outside world (e.g. Mathematica, to some extent)
- On the long term, we suffer from scientific isolation, particularly when the outside world progresses faster than us (e.g. because of economical interest kicking in)
  - R barely used in HEP
  - Our own-made statistics (CLs) has no solid basis in statistics as a science
  - Ignorance and misconception widely diffuse (e.g. are we really frequentists? No, we are not)

PhysStat for RunI

- When the LHC Run I was a bout to start, a few people decided that the *entire* HEP community needed a more solid approach to statistics
  - learn about what other fields do
  - learn about modern research lines
  - start to pay attention to research in statistics
    - e.g. abandoning a home-made dictionary, in favour of the nomenclature developed outside HEP
    - e.g. abandoning non justified shortcuts to known problem (e.g. CLs)

DS@LHC for RunII

- When LHC Run II was about to start, a few other people decided (us) decided that the *entire* HEP community should have taken a similar step with respect to Data Science
  - learn about what other fields do
  - learn about modern research lines
  - start to pay attention to research in data science
    - e.g. abandoning a home-made dictionary, in favour of the nomenclature developed outside HEP
    - e.g. trying to go beyond the established framework of supervised machine learning
    - e.g. approaching this frenetically developing science for applications beyond

Data Science @ LHC 2015

Bridging High-Energy Physics and Machine Learning communities

9 - 13 November 2015, CERN

Local Organising Committee

- Xabier Cid (CERN)
- Gilles Louppe (CERN)
- Michelangelo Mangano (CERN)
- Maurizio Pierini (CERN)
- Jean-Roch Vlimant (Caltech)

Program Committee

- Kyle Cranmer (New York U)
- Cécile Germain (LRI)
- Vladimir Vava Gligorov (CERN)
- Gilles Louppe (CERN)
- Andrew Lowe (Wigner RCP)
- Maurizio Pierini (CERN)
- David Rousseau (LAL-Orsay)
- Maria Spiropulu (Caltech)
- Jean-Roch Vlimant (Caltech)
- Daniel Whiteson (UC Irvine)

International Advisory Committee

- Roger Barlow (Huddersfield U)
- Tommaso Dorigo (INFN-Padova)
- Ian Fisk (Simons Foundation)
- Maria Girone (CERN)
- Eilam Gross (Weizmann)
- Balázs Kégl (LAL-Orsay)
- Constantin Loizides (LBNL)
- Stuart Russell (UC Berkeley)
- Victoria Stodden (UI Urbana-Champaign)
- Max Welling (Amsterdam U)

sponsored by

LHC Physics Center at CERN: http://lpcc.web.cern.ch
Fermilab National Laboratory: http://fnal.gov
Moore-Sloan Data Science Environment: http://cds.nyu.edu/mooresloan

http://cern.ch/DataScienceLHC2015

DS@LHC for RunII

International Advisory Committee

- Roger Barlow (Huddersfield U)
- Tommaso Dorigo (INFN-Padova)
- Ian Fisk (Simons Foundation)
- Maria Girone (CERN)
- Eilam Gross (Weizmann)
- Balázs Kégl (LAL-Orsay)
- Constantin Loizides (LBNL)
- Stuart Russell (UC Berkeley)
- Victoria Stodden (UI Urbana-Champaign)
- Max Welling (Amsterdam U)

Local Organising Committee

- Xabier Cid (CERN)
- Gilles Louppe (CERN)
- Michelangelo Mangano (CERN)
- Maurizio Pierini (CERN)
- Jean-Roch Vlimant (Caltech)

Program Committee

- Kyle Cranmer (New York U)
- Cécile Germain (LRI)
- Vladimir Vava Gligorov (CERN)
- Gilles Louppe (CERN)
- Andrew Lowe (Wigner RCP)
- Maurizio Pierini (CERN)
- David Rousseau (LAL-Orsay)
- Maria Spiropulu (Caltech)
- Jean-Roch Vlimant (Caltech)
- Daniel Whiteson (UC Irvine)

The Program

	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY
TALKS	General Intro DS in HEP ATLAS CMS	ABC / matrix elements / approximate likelihood…	Alice / Deep Learning	LHCb / Real Time data analysis	**Special round table on LHC & Open Data**
TUTORIALS	TMVA	matrix element methods	Deep Learning	SciKit	cuDDN by nVidia

“special” guests

- A few special talks from people working in other fields
  - Ellie Dobson [industry, previously in ATLAS]
  - Gregory Dobler [CUSP]
  - Demis Hassabis [Deep Mind]
  - Christian Mueller [Simons Foundation]
- (as much as we can) a 5-minutes “response” to talks foreseen
  - a physicists to answer a data scientist, and viceversa

Where do we want to go?

- As for PhysStat (and even more, if we manage) we would like to establish a new series of workshops/conferences
  - A bridge between two communities
  - A fixed appointment for our community to do our homework
  - A source of new inputs / new approaches (e.g. not getting stuck with the next NN/BDT/etc for 10 years)
- We are already working on the second appointment
  - Next year, in the US [looking for expressions of interest]
  - Trying to minimise conflict with LHC hot moments (Spring or Fall)

Where do we want to go?

- The first workshop is meant to be a handshake
  - try to cover many topics at once
  - cannot go in the details of each aspect
- Next workshops might be more specific. E.g.
  - concentrate on one method + possible applications
  - concentrate on one problem + possible solutions
- The topic to choose will depend on how the workshop goes

The goal for this workshop

- So far, the goal was to generate and concentrate a diffuse interest of the LHC communities
  - Went beyond our expectations: 200 people registered, many more will connect from vydio (we saturated the capability of the Filtration Plant)
- During the workshop, we would like to get two lists
  - Problems we have / solutions they can offer
- Based on that, there will be some homework to do, in view of the next appointment

How do we do our homework?

- *Within the experiment*
  - every LHC experiment already developed some internal forum to discuss advanced data science
- *Within the Lab*
  - e.g. CERN is engaged to Yandex through the OpenLab project
  - But individual OpenLab projects mainly live inside a given experimental collaboration
- ***Across experiments***
  - It is time to do R&D in the spirit of a no-border community
  - This is why we look fwd to the activity of this group as an established opportunity for people to work together, regardless of their affiliation & membership cards