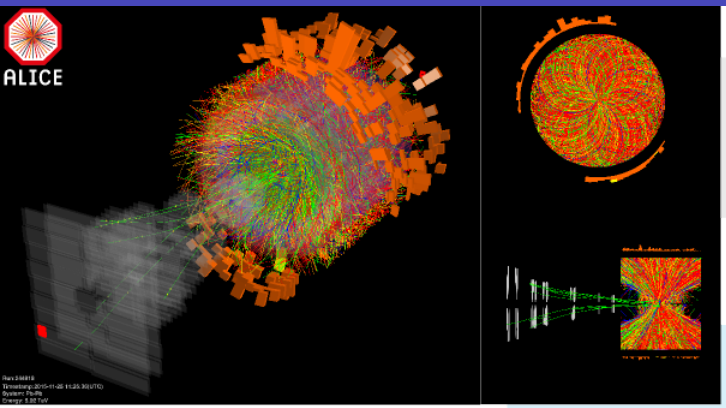
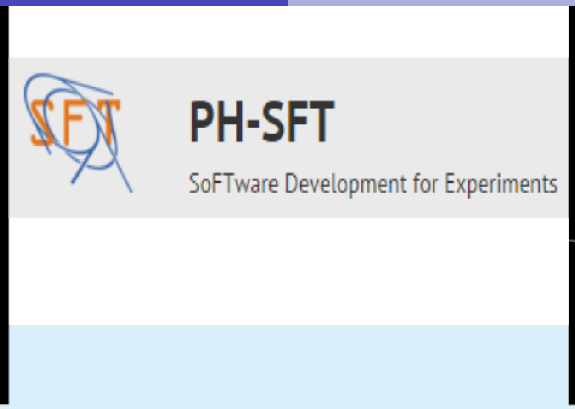
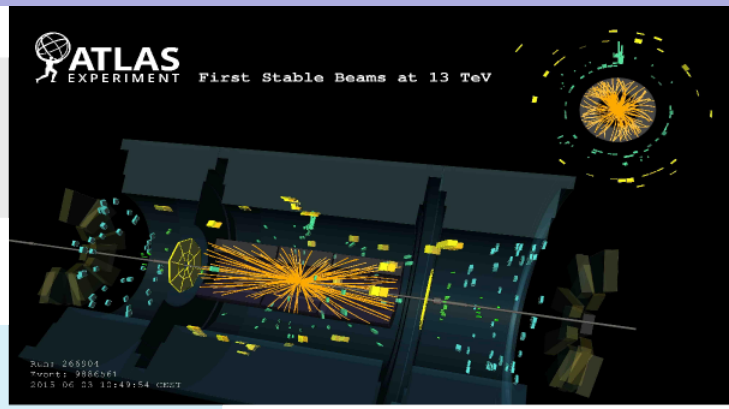
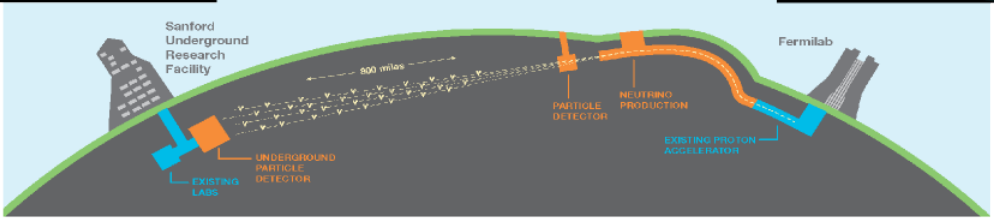




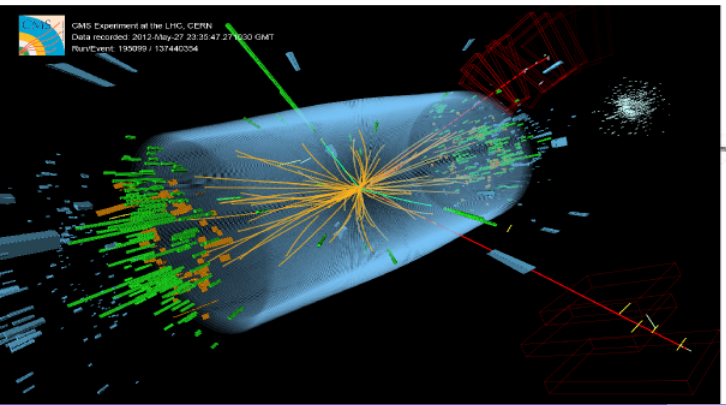


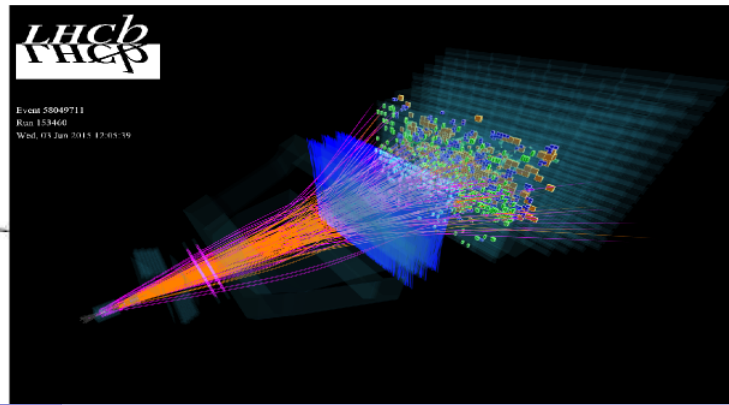


The Inter-experimental Machine Learning Group

Michele Floris, Sergei Gleyzer, Steven Schramm, Tim Head

On behalf of the IML group





Outline

- What is IML?
- Activities
- Format
- Topics
- Future outlook

What is IML

- Inter-experimental Machine Learning Working Group
  - Founded mid-2015
  - Started with a community effort to upgrade and modernize ML methods and tools in HEP
  - Rapidly grew into a mature group endorsed by all LHC experiments in 2016
  - IML website: iml.cern.ch
    - News, Activities, Tasks, Software, Meetings, Forum

IML Format

- Monthly meetings around an ML topic of interest in HEP
  - Work on-going in between meetings
  - Easy to see what others are working on, share ideas, discuss methods and tools useful for all HEP experiments
    - Extremely beneficial to participants
    - Strong and continuously growing community
    - Continuous year-round effort

Topics

- Modernization of HEP ML Tools (1,2)
- HEP-ML Challenges (1)
- ML Software and Tools in HEP (1,2)
- Deep Learning (1)
- Function Estimation with Regression
- Anomaly Detection
- Unsupervised Learning
- Multiclass Classification

Meeting Example

15:00 → 15:10 News and updates 10m

Speakers: Michele Floris (CERN), Steven Randolph Schramm (Universite de Geneve (CH)), Tim Head (Ecole Polytechnique Federale de Lausanne (CH)), Sergei Gleyzer (University of Florida (US))

20160414_IML_ne...

15:10 → 15:30 Bayesian neural networks and general remarks about IML 20m

Speaker: Harry Prosper (Florida State University (US))

IML_Prosper_14Ap...

15:35 → 15:55 Deep learning and NN overview, keras tutorial 20m

Speaker: Michela Paganini (Yale University (US))

NNinKeras_MPaga...

16:00 → 16:20 Gradient reversal layers in deep neural networks for transfer learning 20m

Speaker: Paul Seyfert (Universita & INFN, Milano-Bicocca (IT))

pseyfert.pdf

16:25 → 16:45 Jet images in delphes 20m

Speaker: Daniel Hay Guest (University of California Irvine (US))

dguest_dl.pdf

Non-LHC

16:50 → 17:10 Deep Learning at N0vA 20m

Speakers: Adam Jude Aurisano , Adam Aurisano (Unknown) , Adam Aurisano (University of Cincinnati)

aurisano_CNNid_L...

17:15 → 17:35 Deep Learning Event Reconstruction, From LArTPC to ATLAS Calorimeter 20m

Speaker: Amir Farbin (University of Texas at Arlington (US))

IML-DL.pdf

Software Example

- A community effort identified places where our tools needed improvements
  - Significant work undertaken by IML and the software group to address that challenge (a lot already complete, some in progress)
  - Continuous efforts in evaluation of new ideas in ML: methods, tools, and practices
    - Providing a common ground for testing/evaluation
    - No need for duplication of efforts
  - Significant effort in providing interfaces to interesting and valuable ML tools
    - Bringing them into our ecosystem
    - Building on the work of others

MEM+ML Example

- Last year a potentially interesting new research area identified connecting Matrix Element Methods (MEM) with ML
  - High domain knowledge area in HEP+ML
  - Presented as challenge idea in December IML meeting
  - Discussed in EPlanet Mini-Course in Brazil (SG/LM Nov-Dec. 2015)
  - Participating HEP faculty with significant MEM experience begin active IML involvement with a strong effort in this direction
    - also contributing to relevant ML software work and tasks (deep learning)
    - receiving expert ML help in this area
  - Many similar stories, examples
    - sharing of ideas, tools, solutions leads to progress
  - IML is for the community and by the community
    - A high number of IML volunteers on variety of tasks especially software/tools/interfaces/applied HEP ML

Future outlook of the IML

- One important IML aspect is to facilitate inter-experimental communication on ML
- The IML has now been formally recognized by all LHC experiments
- By also becoming an LPCC group, there are many benefits
  - Access to funds and logistical support
  - Meetings, coffee breaks, workshops, CERN access for non-members, ...
- The IML will continue to involve all interested HEP experiments
  - We've already had presentations from DUNE, NOνA, and ML experts
  - Other experiments have also expressed interest

Support for HEP-ML software

- Another IML aspect: identify critical ML software packages for HEP
- Working closely and in coordination with CERN software group and experiments
  - Can ensure that ML packages meet good software/code standards and are supported long-term
  - When needed, support is possible to bring promising new developments to fruition

Tutorials on ML software

- Another way to increase software support: have more interested users
  - Some of those users can become developers
- It's always good to attract more developers to the field
- The IML will continue to provide tutorials on HEP-ML packages
  - Some active IML members have already given tutorials in the past
  - We hope to do this more often and create an open database of tutorials and ML educational materials

Outreach and Education

- ML Schools, Workshops, Courses, Tutorials
  - 2 ML software tutorials at DS@LHC 2015
  - Eplanet Course + 5 Tutorials in Statistics and ML (Brazil) Nov-Dec. 2015
  - ALICE Workshop in Statistics and ML (lecture and tutorial, May 2016)
  - ESI ML Lectures/Tutorials June 2016
  - Related schools and Conference Sessions:
    - MLHEP Summer School July 2016
    - QCHS Statistics and ML Session Aug 2016

Increasing collaboration with ML experts

- It is important to increase discussion between ML and HEP
- Example of a common scenario:
  - HEP expert has a problem, but doesn't know which ML tool to use
  - ML expert has lots of tools, but doesn't understand the HEP problem
- The IML addresses this by bringing HEP+ML experts together

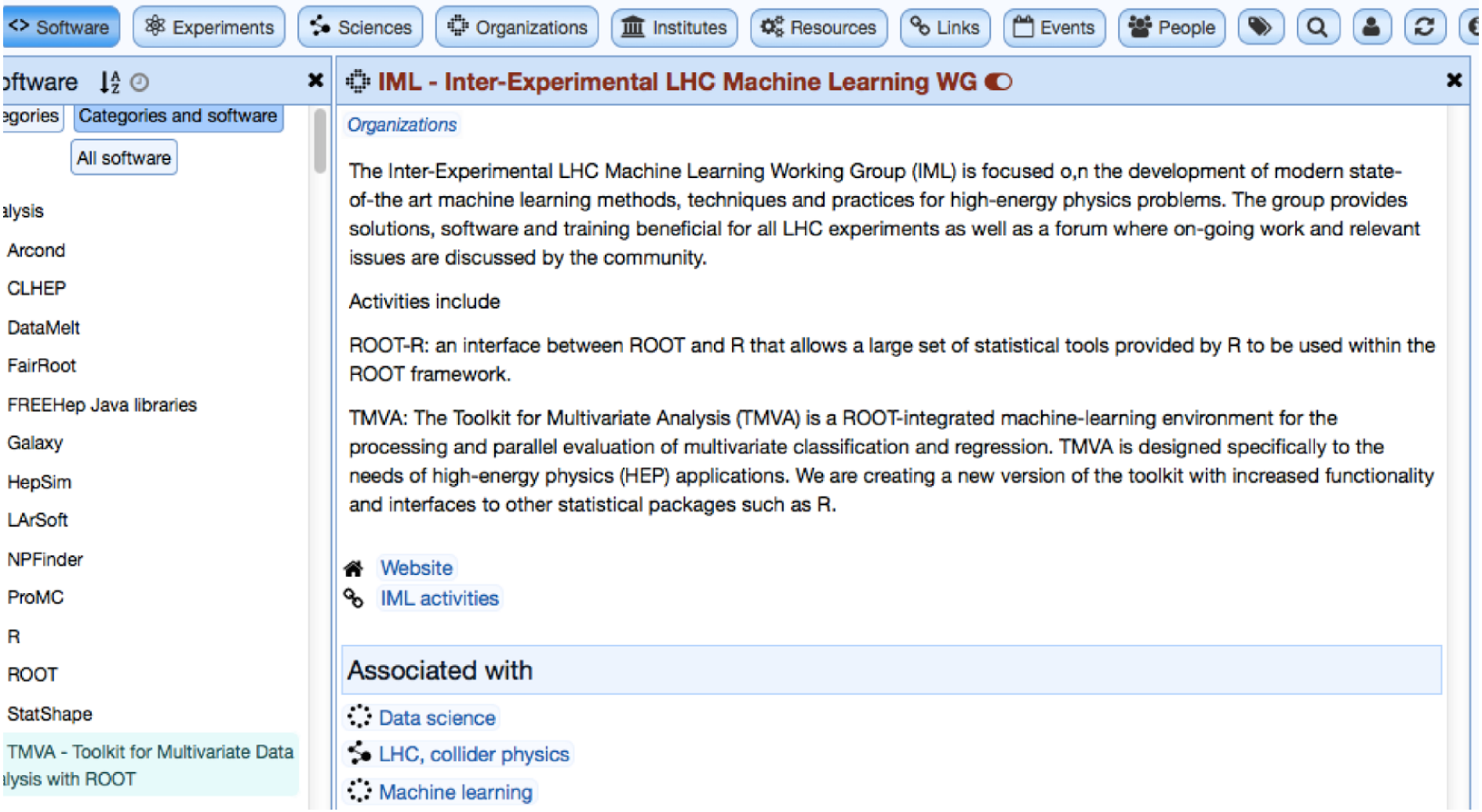
Data Science @ LHC workshop

- The recent DS@LHC workshop was a great success
  - Great contributions from many experiments
  - Lots of input directly from the ML community
- The IML meetings and DS@LHC workshop are mutually beneficial
  - DS@LHC: a *workshop* brings together experts, lots of great discussions
  - IML: a *working group* with regular meetings, updates, and similar

Upcoming meetings

- Next IML meeting: *function estimation with regression*
  - May 17 at 15:00 in the CERN main auditorium (or remotely on vidyo)
- Ideas for future meetings (alphabetically):
  - Anomaly detection
  - Benchmark datasets
  - Multi-class classification
  - Unsupervised methods
  - Suggestions of other topics are welcome!

IML and HSF


<> Software
Experiments
Sciences
Organizations
Institutes
Resources
Links
Events
People
oftware
egories
Categories and software
All software
alysis
Arcond
CLHEP
DataMelt
FairRoot
FREEHep Java libraries
Galaxy
HepSim
LArSoft
NPFinder
ProMC
R
ROOT
StatShape
TMVA - Toolkit for Multivariate Data
alysis with ROOT
IML - Inter-Experimental LHC Machine Learning WG
Organizations
The Inter-Experimental LHC Machine Learning Working Group (IML) is focused o,n the development of modern state-of-the art machine learning methods, techniques and practices for high-energy physics problems. The group provides solutions, software and training beneficial for all LHC experiments as well as a forum where on-going work and relevant issues are discussed by the community.
Activities include
ROOT-R: an interface between ROOT and R that allows a large set of statistical tools provided by R to be used within the ROOT framework.
TMVA: The Toolkit for Multivariate Analysis (TMVA) is a ROOT-integrated machine-learning environment for the processing and parallel evaluation of multivariate classification and regression. TMVA is designed specifically to the needs of high-energy physics (HEP) applications. We are creating a new version of the toolkit with increased functionality and interfaces to other statistical packages such as R.
Website
IML activities
Associated with
Data science
LHC, collider physics
Machine learning

IML and HSF

- During last ROOT workshop in Saas-Fee the idea that IML discussion forum can be used by HSF for ML discussions was proposed (P. Mato, S. Gleyzer, L. Moneta)
  - Great idea that should possibly get more advertisement
  - HSF Open-Source theme resonates with IML
    - IML strives for inclusivity, openness and collaboration among all participants and experiments in this area
    - IML more narrowly focused in Machine Learning in HEP
    - IML participants come from inside and outside HEP

- Some Examples:
  - Google Summer of Code students working on open-source software for HEP community to use (see Lorenzo's talk)
  - ML experts providing input/ideas/advice
  - HEP ML experts often create innovative solutions
    - IML can help raise awareness and bring ideas back to ML community at large

Conclusions

- Inter-Experimental Machine-Learning Working Group is mature and rapidly growing
  - Many ongoing activities
  - Montly meeting format around topics of community interest, plus a lot of work ongoing
  - Significant modernization and upgrade effort of ML-HEP tools and evaluation of new tools + their interfaces
  - Officially endorsed by all LHC experiments
  - More participants are welcome! Anyone interested is welcome to join: visit iml.cern.ch or contact iml.coordinators@cern.ch