Laboratório de Instrumentação e Física Experimental de Partículas

Competence Center - Big Data

4th Informal Meeting – 15th December 2017

ATTENDEES (via Vidyo)

Ana Sofia Nunes, André David, Bruno Galhardo, Celso Franco, David Fernandes, Giles Strong, Guilherme Milhano, Helmut Wolters, Henrique Carvalho, Liliana Apolinário, Marcin Stolarski, Nuno Castro, Tiago Vale

NOTES

- **Agenda:** https://indico.lip.pt/indico/conferenceDisplay.py?confId=337
- **Introduction:**
  - Plans for the future of the competence center. Please send your feedback via email.
  - Dedicated computing resources for machine learning at LIP depend, of course, on the available funding, but the different options must be explored.
  - Survey of tools and interests in big data and machine learning at LIP: please keep updating it at https://docs.google.com/document/d/1Cd0vLKFSb886OuMb583Vo9YPdS_G_8JL9q8N0HHPaN4/edit?usp=sharing
  - Status of the organization of the LIP School and Workshop on Data Sciences (Lisbon, 12-16 March 2018).
- **André David: use of multivariate techniques in the H → $\gamma\gamma$ analysis in CMS**
  - Advanced analysis techniques vs. black boxes.
  - Higgs decays: $\gamma\gamma$ channel is the most sensitive for low mass Higgs; not an obvious choice.
  - Categorization of events is crucial to increase the sensitivity.

- From cut based to MVA: gain equivalent to ~50% more integrated luminosity.
- Intricate analysis anatomy.
- Mass resolution deconstruction: no longitudinal segmentation in CMS EM calorimeter imposes a challenge - angular accuracy is critical to get the narrowest possible mass peak.
- MC to data corrections.
- The black-box fear: some standard sanity checks being done - questionnaire by the CMS statistics forum.
- Primary vertex reconstruction: BDT used; validated with Z → μμ events.
- Di-photon classification: training variables include photon ID, event kinematics, right vertex probability and estimate mass resolution.
- Modelling of data using MC: important to test individual analysis, but also variables. For the $\gamma\gamma$ analysis, the situation is somehow simpler since MVA is not used as final discriminant, but rather as an auxiliary way of categorizing the events.
- The figure of merit for discovery is the uncertainty of the signal strength. For the current phase, targeting precision measurement categorization become even more important, since it allows to disentangle e.g. the different production mechanisms. MVA play a crucial role here.
- Chaining MVAs allowed to unblacken the box quite a lot.
- Reinterpretation in MVA analysis. Recent proposal on the use of MVA for data unfolding: https://arxiv.org/pdf/1712.01814.pdf.

NEXT MEETING

Next meeting will be organized in January/February in date to be defined. Please volunteer to organize the journal club.