Precision Machine Learning for High Energy Physics

Luca Silvestrini
INFN, Rome

Proposal of a new activity in CSN4

INFN Istituto Nazionale di Fisica Nucleare

The Team

- Genova
  - **R. Torre**, S. Marzani (Th)
  - A. Coccaro, F. Parodi, C. Schiavi, F. Sforza (Exp)
- Milano
  - S. Carrazza, S. Forte, M. Zaro (Th)
- Roma
  - L. Silvestrini (Th)
  - S. Giagu, V. Ippolito, C. Mancini Terracciano, A. Messina, S. Rosati, F. Safai Teherani, C. Voena (Exp)

The Problem

- High energy physics is entering a "precision era" with HL-LHC and super KEK-B
- Such abundance of data requires a huge computational effort:
  - theoretical predictions must be refined
  - more MC events must be generated, with higher theoretical accuracy
  - global fits in general frameworks (e.g. SMEFT) require evaluation and distribution of complicated likelihoods
  - flexible and unbiased anomaly detection techniques needed to fully exploit data

The Problem

- Machine Learning techniques very useful (and extensively used) in both
  - classification: tagging, clustering, anomaly detection, etc.
  - generation: MC events, detector simulations, data augmentation, etc.
- The precision era of HEP requires precision ML, with quantitative estimates of uncertainties and explainability

The Work Plan

1) Theory predictions

2) Data acquisition

3) Data analysis and statistical inference

4) Presentation, distribution and preservation of results

5) General ML tools and other applications

Theory Predictions

- Use generative networks to speed up
  - phase space integration in matrix element calculation
  - showering and hadronization
  - detector response simulation
  - MonteCarlo and MCMC sampling of theoretical and phenomenological pdfs in global Bayesian analyses

Data Acquisition

- Implementation of ML methods for online triggering and tagging
- Implementation of fast inference methods for online anomaly detection and data quality monitoring
- Implementation of ML methods for offline jet reconstruction and object tagging, possibly with the associated uncertainty

Data Analysis and Statistical Inference

- Statistical inference in the SMEFT with parametrized classifiers to exploit the theoretical knowledge about the dependence of observables on SMEFT parameters
- Likelihood-free inference using density estimation
- Bayesian networks for systematic uncertainties

Presentation, Distribution and Preservation of Results

- Use ML approximation methods to encode and distribute experimental or phenomenological likelihoods with O(100-1000) parameters
- Extend method to cases in which only a sample of the pdf is available
- Provide robust metric for accuracy and generalization properties

General ML Tools and Other Applications

- Develop new sampling and MC integration techniques
- Develop new metrics for two-sample tests relevant for physical applications
- Implement physics informed algorithms that exploit symmetries or properties of data
- Study general properties and scalability of density estimation and generative models performance
- Emulate time-consuming theory computations to accelerate nuclear interaction methods of interest for medical applications