Contribution ID: **50**   Type: **Poster**

Conditional normalizing flows for correcting simulations

Thursday 1 February 2024 15:50 (5 minutes)

Simulated events are key ingredients for almost all high-energy physics analyses. However, imperfections in the configuration of the simulation often result in mis-modelling and lead to discrepancies between data and simulation. Such mis-modelling often must be taken into account by correction factors accompanied by systematical uncertainties, that can compromise the sensitivity of measurements and searches.

To address this issue, we propose to use normalizing flows, a powerful technique for learning the underlying distributions of input data. We employ a conditional normalizing flow that utilizes as conditions kinematic variables together with a boolean that differentiates between simulation and data. By training the flows with both simulation and data and mapping both distributions to the same latent space, the flow can learn both underlying distributions and map between them using the latent space as an intermediary. Thus, one can map simulated events to the latent space, and then swap the conditional boolean. Due to the invertibility of the flows, one can then map from the latent space to the data space, yielding the corrected values for the variables. The most important innovation of this method is that a single flow is sufficient to morph a multidimensional distribution into another.

We demonstrate that the proposed architecture can transform simulated events into data using a toy example. The toy distributions are inspired by physical distributions, where the variables are generated conditioned on kinematic variables, and the resulting distributions exhibit correlations between themselves. These correlations and kinematic distributions differ for simulation and data distributions. The distributions include non-continuous functions, which are handled with suitable transformations. We assess the quality of the corrections by training a classifier on the toy data and the corrected simulated events.

Primary authors: DAUMANN, Caio Cesar (Rheinisch Westfaelische Tech. Hoch. (DE)); VALSECCHI, Davide (ETH Zurich (CH)); SPAEH, Jan Lukas (Rheinisch Westfaelische Tech. Hoch. (DE)); ERDMANN, Johannes (RWTH Aachen University); GALLI, Massimiliano (ETH Zurich (CH)); DONEGA, Mauro (E)

Presenter: DAUMANN, Caio Cesar (Rheinisch Westfaelische Tech. Hoch. (DE))

Session Classification: Poster Session

Track Classification: 2 ML for analysis : event classification, statistical analysis and inference, including anomaly detection