2. LHCb Masterclass, with Keras

Author(s)

Name	Institution	Mail Address	Social Contacts
Lucio Anderlini	INFN Sezione di Firenze	Lucio.Anderlini@fi.infn.it	Hangouts: l.anderlini@gmail.com
Matteo Barbetti	Università di Firenze	Matteo.Barbetti@fi.infn.it	N/A

How to Obtain Support

Mail	Lucio.Anderlini@fi.infn.it
Social	Hangouts: l.anderlini
Jira	N/A

General Information

ML/DL Technologies	Statistical Learning; Forward Neural Networks
Science Fields	High Energy Physics
Difficulty	Introductory
Language	English
Type	fully annotated

Software and Tools

Programming Language	Python
ML Toolset	Keras + Tensorflow
Additional libraries	uproot
Suggested Environments	INFN-Cloud VM, bare Linux Node, Google CoLab

Needed datasets

Data Creator	LHCb Experiment
Data Type	2011 data
Data Size	1 GB

Data Source	CERN OpenData

Short Description of the Use Case

For the outreach programme *LHCb Masterclass* students from secondary schools are invited to analyze a sample of D0 K pi+ decays as collected from the LHCb experiment to measure the lifetime of the D0 meson.

The data used for this exercise are public and can be obtained from the *Open Data* portal of CERN.

In this tutorial we repeat the analysis designed for the LHCb Masterclass, using *Python* and *ROOT* in order to show how the most common operations in data analysis can be performed within such a framework.

We will take the opportunity to apply some *machine learning*, this is not part of the original excercise, but it is worth to include an example on how to use *Keras* and *Tensorflow* to separate signal and background. This is not a lecture on *machine learning*: several basic and important aspect of a machine learning problem are ignored here, for example we do not split the data in *training* and *test* samples. From a software perspective, it should be trivial to extend the example to include a more careful treatment of the neural network training and application.

The website of the *LHCb International Masterclass*, where the excercise is shortly explained can be found at this link.

How to execute it

Requirements

To run this exercise you will need python3, tensorflow 1.x and PyROOT for python3.

Download and run the jupyter notebook: https://github.com/landerlini/MLINFN-TutorialNotebooks/blob/master/LHCbMasterclassExplained.ipynb

Contents

With this tutorial, we will introduce the following topics:

1. Download data with jupyter via http
2. Exploring a dataset with pandas
3. Exploring a dataset with matplotlib
4. Obtaining high quality plots with ROOT
5. Modelling the data with RooFit
6. Perform a per-event subtraction of the background using sPlot
7. Training a simple neural network on nTuple data with keras
8. Evaluate the performance of the trained algorithm
9. Apply the neural network to data
10. Studying systematic effects induced by the neural network

ⓘ Presentation made on 27 Jul 2020 : https://agenda.infn.it/event/23648/contributions/118758/attachments/74371/94574/Teaching_Material_on_the_KB.pdf