Machine learning for the analysis of low-mass dielectrons in Run II data with ALICE

A. Chauvin[*1,2], *T. Dahms*[1,2], *A. Dashi*[1,2], *O. Vasquez Doce*[1,2], *and I. Vorobyev*[1,2]

[1]Excellence Cluster Universe, Technische Universität München, Garching, Germany; [2]Physik Department, Technische Universität München, Garching, Germany

Dielectron pairs are an experimental tool to investigate the Quark Gluon Plasma (QGP), which is expected to be created during ultra-relativistic heavy-ion collision. The measured electron-positron pairs are created at different stages of the evolution of the hot and dense medium and do not interact strongly with the latter. Hence, dielectron pairs can probe the full space-time evolution of the system, thereby allowing us to investigate the predicted restoration of chiral symmetry as well as the initial temperature of the QGP.

Machine Learning is a statistical tool that allows combining a large number of variables to perform classification and regression analysis known as Multi-Variate Analysis [1]. Combined with complementary analyses, such as overtraining studies and reweithing, a Multi-Variate Analysis can be performed in our analysis e.g. for background rejection and electron identification.

Photon conversions contribute to the background of the dielectron signal. Besides contributing to the combinatorial background, electron pairs from photon conversions contaminate our signal in the very low-mass region (<100MeV). So far a cut on φ_V (the pair orientation relative to the magnetic field direction) [2] is applied. A cut of φ_V <2 rad leads to a signal efficiency of 61% and a background rejection of 99%. The Multi-Variate Analysis allows us to consider a total of 16 track and pair variables which are then combined using Boosted Decision Trees. The optimal cut to apply is found when performing overtraing. This consists of looping over the hyperparameters phase-space to look for the first local maximum of the Receiver Operating Curve of the test sample as shown in Figure 1. This way, one can ensure the selected parametrisation does not depend on the training sample. Thus, for the same background rejection of 99%, a signal efficiency of 89% is achieved by a cut on the newly computed variable.

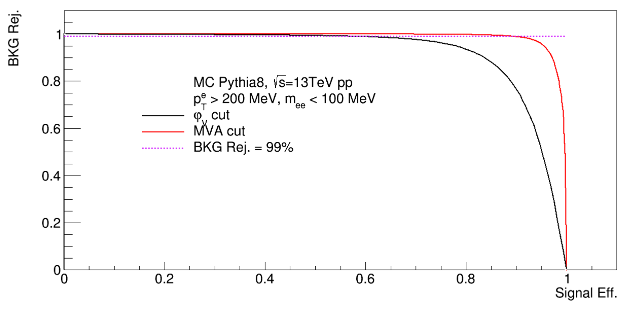

Figure 1: Receiver Operating Characteristic curve for cuts on φ_V and on the MVA output, the background rejection is plotted as a function of the signal efficiency

On top of conversion pairs, MVA can also be used to reject combinatorial background and to perform electron identification. In order to tune Monte-Carlo simulation to match data, reweighing methods [3] can be applied on different type of pairs to be classified. This way, an estimation of the different contribution from the different classes to the spectrum can be evaluated [4]. A MVA relying on Monte-Carlo simulation can be performed for electron identification. In ALICE [5] particle identification is performed using three main detectors: the Inner Tracking System, the Time Projection Chamber and the Time Of Flight. So far, cuts are applied separately on the signals in these different detectors. With the help of reweighing Monte-Carlo simulation can be used on top of other track variables to perform identification [6]. In this case on can expect an increase of the efficiency and the purity, which would reduce the statistical and systematic errors.

References

[1] A. Hocker *et al.*, PoS ACAT (2007) 040 [physics/0703039 [PHYSICS]].

[2] T. Dahms, arXiv:0810.3040 [nucl-ex].

[3] A. Rogozhnikov, J. Phys. Conf. Ser. **762** (2016) no.1, 012036 doi:10.1088/1742-6596/762/1/012036 [arXiv:1608.05806 [physics.data-an]].

[4] S. Lehner, ALICE-ANA-3478.

[5] K. Aamodt *et al.* [ALICE Collaboration], JINST **3** (2008) S08002. doi:10.1088/1748-0221/3/08/S08002

[6] A. Capon, ALICE-ANA-3486.

* alex.chauvin@cern.ch