Development of a method for fast, easy and optimized tuning of selection cuts

Modification of the Fisher Linear Discriminant Analysis for use in the low signal-to-noise environments

J. Faivre*

INFN sezione di Padova, Padova, Italy

Received: 30 June 2006 / Published online: 3 August 2006 – © Springer-Verlag / Società Italiana di Fisica 2006

Abstract. In situations where the signal of the analysed particle is tangled up with orders of magnitude more background, its analysis may benefit from the use of a pattern classification method to discriminate the signal out of the background candidates. We present and explain the basic Linear Discriminant Analysis and the modifications brought – the use of cascaded cuts and of a locally optimized criterion – to adapt it to the conditions encountered in the field of heavy ion Physics. We show that this optimized multicut Linear Discriminant Analysis has a higher performance than classical selection cuts and provides a very fast and easy selection cut optimization.

1 Introduction

Particle search in a collision event consists in discriminating the signal (what is wanted) and the background, or noise (fake candidates). This pattern classification is achieved through the measurement of several characteristics for each candidate, herein called cut variables or observables, and the discrimination relies on the fact that the probability distributions of these characteristics are different for the signal and the background.

The simplest method consists in applying a “straight selection cut” on each of these variables separately, as shown in the top of Fig. 4. We will refer to this approach as classical analysis or classical cuts. Because these – numerous – cut values are considered as independent parameters while those variables are generally correlated, this method may be very long to tune and usually provides an improvable discrimination. In this article, we will describe the adaptation of Fisher Linear Discriminant Analysis (Fisher-LDA), a pattern classification method widely used in data processing, to the extreme signal-to-noise conditions of the relativistic heavy ion collisions, and show its advantages over the “classical analysis”.

The first section explains why such methods are needed for heavy ion Physics, and gives examples. The second section is a short introduction to pattern classification. Fisher-LDA and its modifications are presented in the third and fourth sections. Finally, the last section explains how the final multi-variable cut is tuned and used in practice, and shows some results. Duda et al. and Faisan [1,2] have helped writing Sects. 3.1 and 4.2.

The method described in this paper has been implemented as a plug-and-play C++ class. Its source code and documentation are available upon request to the author.

2 Low signal-to-noise environments

As an example of low signal-to-noise environment, we introduce here briefly the context which led us to use the LDA method. More information about (ultra-)relativistic heavy-ion collisions can be found elsewhere, e.g. [3–5].

2.1 Relativistic heavy-ion collisions

Lattice-QCD predicts that when the energy density of a strongly interacting system of hadrons is large enough, matter should undergo a phase transition from a hadronic state to a quark-gluon plasma (QGP), in which the degrees of freedom are partonic. Parton deconfinement is made possible in such a medium by the strong force screening resulting from the high density of colour charges, similarly to the Debye screening in an electromagnetic plasma.
In this purpose, (ultra-)relativistic heavy ion (Pb–Pb, Au–Au) collisions are made, with $\sqrt{s_{NN}}$ currently ranging from a few GeV (AGS) to 200 GeV (RHIC), and up to 5.5 TeV when LHC starts to run.

Some of the probes of the QGP are based on strangeness and charm, absent in the initial state of the collision. Hadrons containing those quarks decay weakly, which enables an identification up to transverse momenta in the pQCD domain.

Those particles are best studied by reconstructing topologically their secondary decay vertex. This can be achieved in the detectors STAR at RHIC [6], and ALICE at LHC [7], thanks to the tracking subdetectors: in both cases, a time projection chamber (TPC) surrounding several layers of silicon detectors makes possible the reconstruction of the silicon detectors makes possible the reconstruction of the trajectory of the charged particles and their extrapolation towards the primary vertex.

### 2.2 Cut variables

This paragraph gives examples of cut variables which can be used for discriminating between the signal and the background (fortuitous associations of tracks) in the case of a $\Lambda^0 = uds \rightarrow p\pi^-$ analysis by topological reconstruction.

A weak decay is characterized by a sizeable decay length ($ct$ of a hundred microns for charm decays, of a couple of centimeters for strange decays). The reconstruction of a neutral particle decay (V0 vertex) is made by examining all combinations of pairs of opposite charge tracks, and filtering out those (background) which have a geometry incompatible with that of a real particle (signal).

In reality, because of the finite resolution of the detectors, the reconstruction is not perfect and real particles and a significant fraction of the background have a similar geometry. This makes the discrimination challenging, and achievable only statistically: the candidates selected as signal are mostly signal, those which are filtered out are mostly background. The proportion of signal kept or rejected by the selection process can be estimated by simulation studies for instance.

The projection in two dimensions of the geometry of a V0 vertex is shown in Fig. 1. The charged tracks are bent by the magnetic field (here perpendicular to the figure plane), and because the reconstruction is imperfect, the tracks of the two decay daughters do not cross and the trajectory of the reconstructed parent particle does not meet the primary vertex.

The decay length, the distances of closest approach between the tracks, or between a track and the primary vertex, constitute geometrical variables which can be used to discriminate the background and the signal. Most of these variables are correlated, e.g. the distance of closest approach between a daughter track and the primary vertex is correlated with the A decay length.

The cosine of the decay angle ($\cos \theta^*$) is also often used to eliminate the background: the distribution of this variable shows strong peaks at $-1$ and $+1$ for the background. Examples of other cut variables may be found in [8], and in [9, 10] for other analyses.

### 2.3 Examples of signal-to-noise ratios

Heavy-ion collisions make topological reconstruction of the weak decays a challenging task, because of the high charged track density (multiplicity) in the detectors. The amount of background for a 2-particle decay scales with the square of this multiplicity, while for of a 3-particle decay it scales with its cube.

For the case of the $\Omega^− = sss \rightarrow \Lambda^0K^−$ in STAR’s central collisions, the yield of about 0.6 $\Omega + \overline{\Omega}$ per event [11] and the multiplicity of more than 3000 tracks give an initial signal-to-noise ratio only slightly above $10^{-10}$. At the reconstruction stage, loose cuts are applied to reduce the computing time and the disk space taken by the storage. While these cuts remove 99.99% of the combinatorics, the signal-to-noise ratio is still as low as $10^{-6}$.

For the $D^0 = c\pi \rightarrow K^−\pi^+$ in ALICE, the initial signal-to-noise ratio is of the order of $10^{-8}$ [12]. Although this is higher in value than for the $\Omega$, the fact that the signal and background distributions of the geometrical variables differ more in the case of the $\Omega$ than in that of the $D^0$ makes the latter more difficult to reconstruct than the $\Omega$.

Analyses in such extreme conditions, also encountered in the fields of top quark analysis or Higgs search, benefit from the advantages brought by the pattern classification methods. Other fields – industry, health, image processing in general – do not deal with such situations, but rather with poor training statistics and high numbers of observables and/or classes. The methods created for their needs therefore do not meet ours, which made necessary the development of a method adapted to our conditions.

---

1. Here, not calculated in an invariant mass window selecting the signal peak. It can therefore not be compared with the numbers given in Fig. 6.

2. Reconstruction of the secondary decay candidates from the tracks, themselves reconstructed from the hits in the detectors.