Advanced Topics in Computational Physics
Master's Degree in Physics

Discrimination of dimuon production events in neutrino interactions in the context of the MINERνA experiment

---

Sebastián Sánchez, Javier Rengifo, Nhell Cerna, y Marvin Ascencio

Physics Section

Pontificia Universidad Católica del Perú

San Miguel, July 4, 2016

Abstract

In the present work, we aim to optimize the discrimination of dimuon production events in neutrino - nuclei interactions in the context of the MINERvA experiment. For this type of events, the main background are general dilepton production events. Our dataset was simulated with the GENIE Neutrino MC v2.8.6, which will also provide us with a simplified parameterization of the MINERvA experiment. The signal discrimination is done over experimentally reconstructable variables and optimized with TMVA [1] package v4.2.0. We finally discuss on the significance of our results for the search of new physics and nuclear structure.

Introduction

The production of lepton pairs induced by neutrino scattering in a Coulumb field of a target nucleus is called neutrino trident production and it offers further possibilities to study neutrino-nuclei interactions. It is described by the general reaction:

$$\nu_\ell(\bar{\nu}_\ell) + N \rightarrow \nu_\ell(\bar{\nu}_\ell) + \ell^+ + \ell^- + N \tag{1}$$

Where N is a nucleus. The case in which $\ell = \mu$ is of most interest because of experimental reasons: high energy neutrino beams rely on meson decay for producing neutrinos, a process for which the muon channel is significantly favored. Also, muon tracks are very clean and thus permit good reconstruction of the neutrino interaction event kinematics.

In standard model reaction [1] can proceed via two interfering channels: charged (W) and neutral (Z) boson exchange (see Fig. 1).

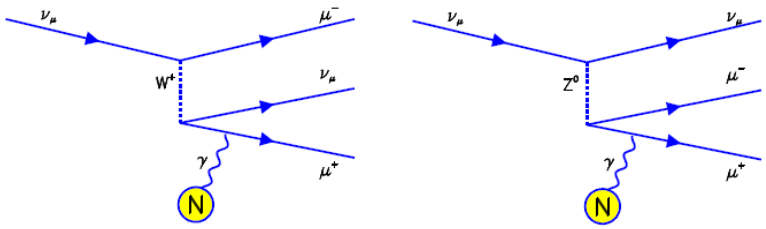


Figure 1: Feynman diagrams for neutrino trident production [2].

Searches for neutrino trident production have been conducted in past experiments using a variety of neutrino energy spectra and targets [2, 3, 4, 5, 6]. The signal tagging relied on demanding a pair of muons with small invariant mass and little hadronic energy deposit near the vertex.

A broader class of events is that of (visible) dilepton production from neutrino-nuclei interactions. These events usually include short lived meson decays or pair production from high energy photons. They constitute the major source of background for neutrino trident production. Considering also that the trident cross section is relatively small, $\sigma_{trident} \sim 10^{-6}\sigma_{CC}$ [7], it is imperative to have reliable methods for discriminating the signal (dimuon events) from background (dilepton

events). This makes Multivariate Analysis (MVA) Methods attarctive as a way to extract as much information as possible from the tipically small samples of dilepton events.

We use the the Toolkit fo MultiVariate Analysis, TMVA v4.2.0 [1], which is a ROOT [8] based package for the training, testing and application of both linear and nonlinear MVA methods. Our data sample is generated with a custom modified event generation script of the GENIE Neutrino Interaction MC [9]. We perform the event generation in the context of the MINERνA experiment by using it's neutrino input flux and a simple parameterization of the detector.

1 Methodology

1.1 Data collection and Processing

In this section we briefly describe the methods used for event generation, preprocessing and multivariate analysis.

HEP interaction event generation can be a time consuming process. For example, in GENIE, one can expect that an inclusive (all the curently implemented interaction models) event sample of one million events take as much as around an hour to be generated in a single core machine running at $2.4GHz$. Fortunately, event generation belongs to the class of trivially parallelizable problems, so the time needed can be cut significantlty in a multicore server.

One may further disable some of the interaction channels if they are not of interest. In our case, however, this is not a viable option, since we need to examin all the porssible channels that would eventually lead to exactly one pair of leptons in the final state. This in turn poses a new problem: storage. As most of the generated events in an inclusive sample don't fulfill this requirement, storing all the events is useless. Then, we custom modified one of the event generation programs in GENIE to compute an inclusive sample and only saves the events that contain specifically a pair of charged leptons in the final state dsregarding flavor. Thus, charged leptons are allowed to come from the first scattering or from the decay of short lived particles, like mesons.

Finally, also motivated by the experimental cuts reported in the literature for trident production searches. All the written and tested code is properly maintained and documented in a GitHub repository (`https://github.com/sfsanche/bazo_proyecto`)

1.2 Multivariate Analysis

In this section we show a brief resume about the methods of TMVA [1] that we use in this work for the analysis of the data. Also we extract some definitions from TMVA web: `http://tmva.sourceforge.net/`

1.2.1 Rectangular cut optimisation

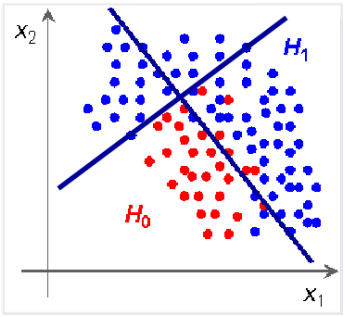
The simplest and most common classifier for selecting signal events from a mixed sample of signal and background events is the application of an ensemble of rectangular cuts on discriminating

variables.

Simplest method: cut in rectangular volume using

$$x_{\text{cut}}(i_{\text{event}}) \in \{0,1\} \quad = \prod_{v \in \{\text{variables}\}} \left(x_v(i_{\text{event}}) \subset \left[x_{v,\min}, x_{v,\max} \right] \right)$$

Cuts usually benefit from prior decorrelation of cut variables



Optimal cuts maximise the signal efficiency at given background efficiency. Other optimisation criteria, such as maximising the signal significance-squared, $S^2/(S+B)$, with S and B being the signal and background yields, then correspond to a particular point in the optimised background-rejection versus signal-efficiency curve. Also there are three optional methods: 1. Monte Carlo generation (option: MC). 2. Fitting using a Genetic Algorithm (option: GA). 3. Fitting using Simulated Annealing (option: SA - still in testing phase).

1.2.2 k-Nearest Neighbour (k-NN) Classifier

The k-nearest neighbour method compares an observed (test) event to reference events from a training data set. It searches for a fixed number of adjacent events, which then define a volume for the metric used. The k-NN classifier has best performance when the boundary that separates signal and background events has irregular features that cannot be easily approximated by parametric learning methods.

1.2.3 H-Matrix discriminant

It discriminates one class (signal) of a feature vector from another (background). The correlated elements of the vector are assumed to be Gaussian distributed, and the inverse of the covariance matrix is the H-Matrix.

Two χ^2 estimators are computed for an event, each one for signal and background, using the estimates for the means and covariance matrices obtained from the training sample. TMVA then uses as normalised analyser for event the ratio: $[\chi^2_s(i) - \chi^2_B(i)]/[\chi^2_s(i) + \chi^2_B(i)]$

1.2.4 Fisher discriminants (linear discriminant analysis)

Event selection is performed in a transformed variable space with zero linear correlations, by distinguishing the mean values of the signal and background distributions. The linear discriminant analysis determines an axis in the (correlated) hyperspace of the input variables such that, when projecting the output classes (signal and background) upon this axis, they are pushed as far as possible away from each other, while events of a same class are confined in a close vicinity. The metric "far apart" and "close vicinity" are determined: the covariance matrix of the discriminating variable space.

1.2.5 Linear discriminant analysis (LD)

The linear discriminant analysis provides data classification using a linear model, where linear refers to the discriminant function. Similar to Method Fisher.

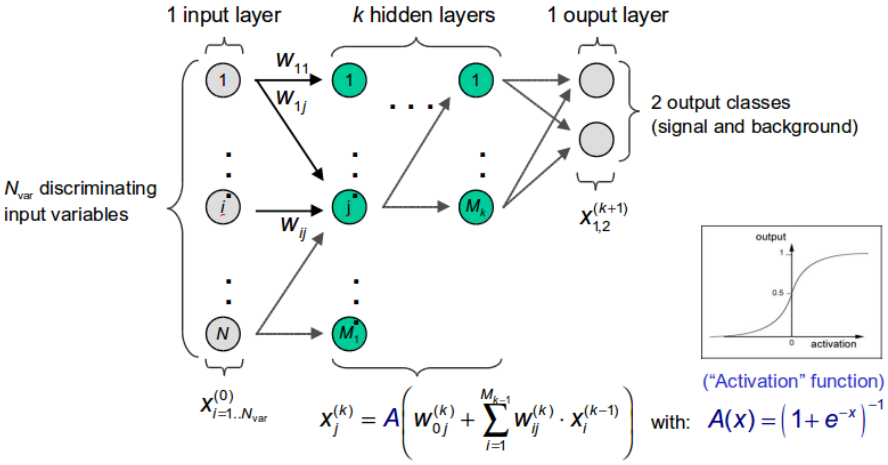
1.2.6 Function discriminant analysis (FDA)

This method provides an intermediate solution to the problem with the aim to solve relatively simple or partially nonlinear problems. The user provides the desired function with adjustable parameters via the configuration option string, and FDA fits the parameters to it, requiring the function value to be as close as possible to the real value (to 1 for signal and 0 for background in classification). Its advantage over the more involved and automatic nonlinear discriminators is the simplicity and transparency of the discrimination expression. A shortcoming is that FDA will underperform for involved problems with complicated, phase space dependent nonlinear correlations.

1.2.7 Artificial Neural Networks (nonlinear discriminant analysis)

An Artificial Neural Network (ANN) is most generally speaking any simulated collection of interconnected neurons, with each neuron producing a certain response at a given set of input signals. By applying an external signal to some (input) neurons the network is put into a defined state that can be measured from the response of one or several (output) neurons.

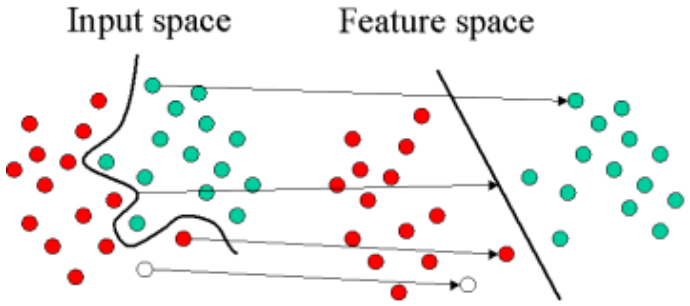
Achieve nonlinear classifier response by "activating" output nodes using nonlinear weights



1.2.8 Support Vector Machine (SVM)

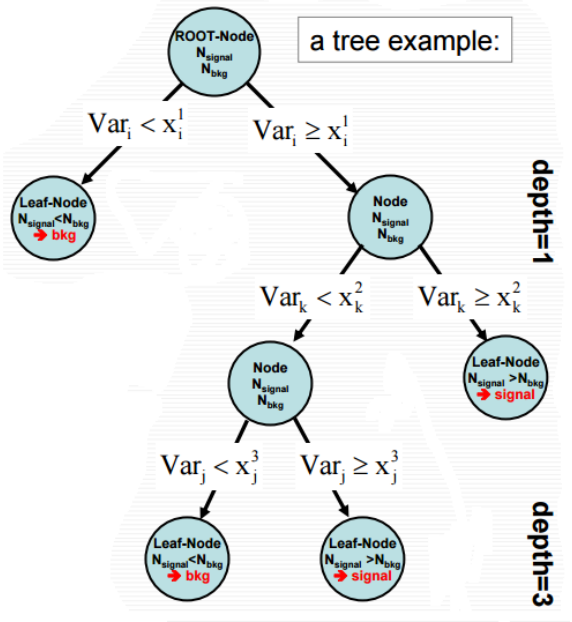
Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. The separation should be lineal or more complicated.

The illustration below shows the basic idea behind Support Vector Machines extract to the web http://www.statsoft.com/Textbook/Support-Vector-Machines



1.2.9 Boosted Decision and Regression Trees

Decision Trees: a sequential application of a tree example: "cuts" which splits the data into nodes, and the final nodes (leaf) classifies an event as signal or background. The devision is stopped once a certain node has reached either a minimum number of events, or a minimum or maximum signal purity.



Boosting: the idea behind the boosting is, that signal events from the training sample, that end up in a background node (and vice versa) are given a larger weight than events that are in the correct leave node. This results in a re-weighed training event sample, with which then a new decision tree can be developed. The boosting can be applied several times and one ends up with a set of decision trees (forest).

Analysis: applying an individual decision tree to a test event results in a classification of the event as either signal or background. For the boosted decision tree selection, an event is successively subjected to the whole set of decision trees and depending on how often it is classified as signal, a "likelihood" estimator is constructed for the event being signal or background. The value of this estimator is the one which is then used to select the events from an event sample, and the cut value on this estimator defines the efficiency and purity of the selection.

1.2.10 Predictive learning via rule ensembles (Rule-Fit)

The discriminator is a linear combinations of base learners called rules. A single rule is essentially a function of a series of cuts. A rule applied on a given event is non-zero only if all cuts in the product are satisfied. RuleFit consists of two main steps: 1. Rules generation. 2. Fit of the rules to the training data, i.e, find the optimum coefficients.

Others Methods that we don't use

We don't use some methods for our analysis because they shown poor results to separate signal and background. This Methods were: Projective likelihood estimator (PDE approach), Multi-dimensional likelihood estimator (PDE range-search approach) and Likelihood estimator using self-adapting phasespace binning (PDE-Foam).

2 Discussion

In this section we show plots that describe the most important features about the output that we obtained using TMVA.

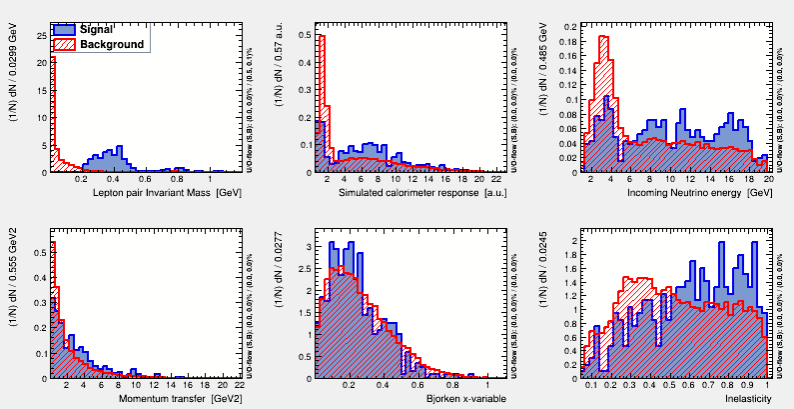
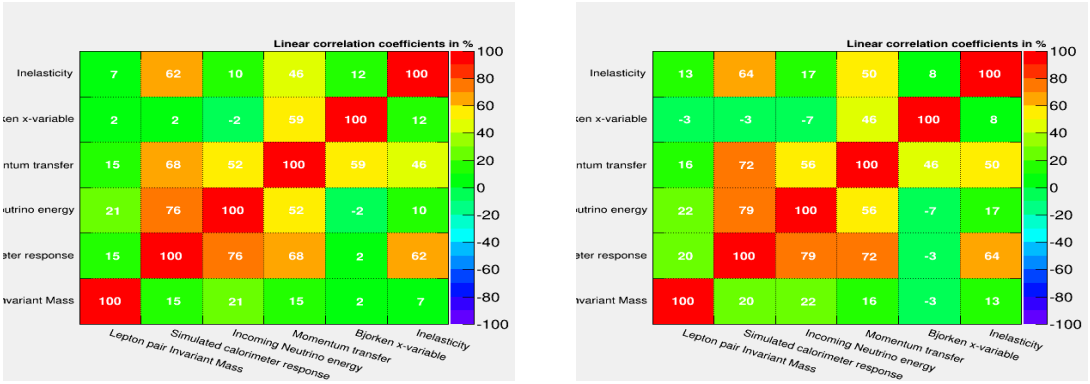


Figure 2: Plots of Input variables (training sample) show histograms of signal in blue and background in red

From figure 2, we can see that the lepton pair invariant mass is the best input variable to separate signal and background, because is posible fix a simple cut ans separate the signal with a little contamination.

Coorelations between input variables are evaluated by TMVA as part of its initial processing. In Fig. 3, the linear correlation coefficients between variables is shown for both signal and background. Most combinations don't exhibit strong correlations, and the ones that do are expectable. For instance, one should expect the energy deposit (and thus the calorimeter response) to increase with increasing inelasticity or momentum transfer in the scattering, as more hadronic final states are likely to be produced. These also depend on the incoming neutrino energy being high enough so that the inelastic (and deeply inelastic) interaction channels' cross sections become significant enough to be dominant over tother processes.



(a) Signal (b) Background

Figure 3: Input variable linear correlation coefficients

As MVA may benefit from prior decorrelation of input variables [1], several decorrelation meth-

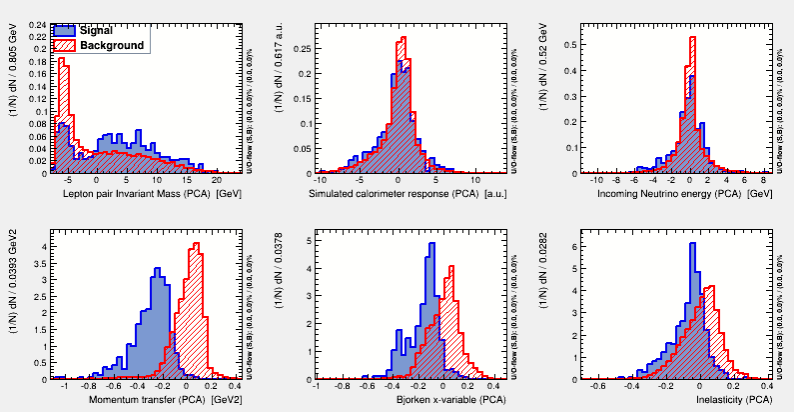


Figure 4: Plots of Input variables (training sample) show histograms of signal in blue and background in red with PCA transformed

ods are tested by TMVA. However, in figure 4 it is shown that it's possible to improve the separation between signal and background for variables like momentum transfer and Bjorken-x, but the distinction worsens for lepton pair invariant mass because the signal gets mixed with background. A simmilar situation happens with other decorrelation transformations (by square root of correlation matrix and gaussianisation).

Another important topic for disscusion in the MVA analysis is the overtraining. Figures **??** show the classifier output distributions for both signal and backgraound and for testing and training samples for three repesentatie methods. These plots show how well is the signal separated from the background. In addition, if overtraining is present, the test and training distributions will differ. This can be evaluated with, e.g. a Kolmogorov Smirmov (KS) test. For example, in figure 5 it is possible to see that KS test value is 0 for signal and 0.132 for background. This indicates overtraining, which is somewhat expectable for BDT considering the high number of degrees of freedom that it comprises.

Also we show for comparison the same plot for KNN, in which the signal and background are extremely separated and the Kolmogorov-Smirnov test is very small (figure 6)

The poorest method to differenciate signal and background is the likelihood classifier; see the significant amount of signal and background overlap in figure 7.

For each MVA method, the optimal cut values are chosen by finding a compromise between signal purity and efficiency. This quantities are evaluaated at varying cut values, for example, for BDT in fig.8. Maximum significance, defined as $S = S/(S + B)$ defines the optimal cut value.

To compare all the evaluated methods in a single plot, we show in figure 9 the ackground rejection vs. signal efficiency. Ideally, a perfect method should reach 100% of both. We can see that MLPBNN, BDT and KNN perform comparatively better than, e.g. likelihood classifiers.

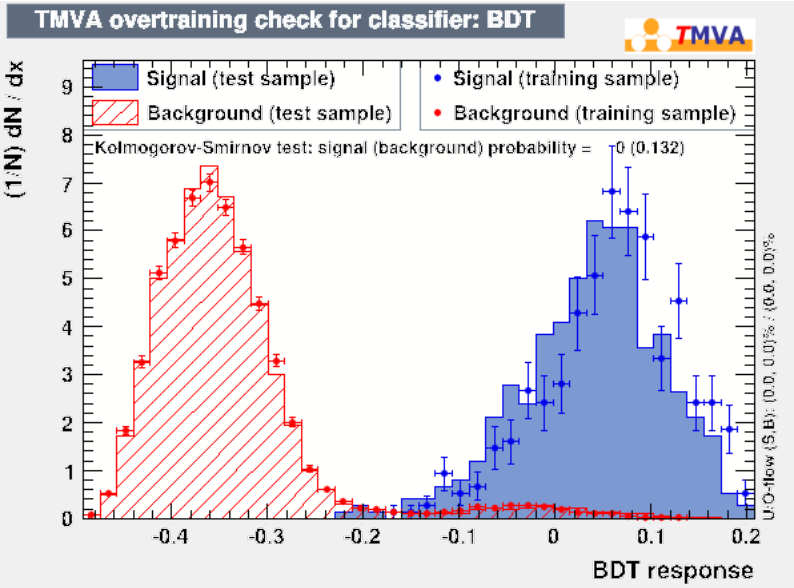


Figure 5: TMVA overtraining check for classifier: BDT

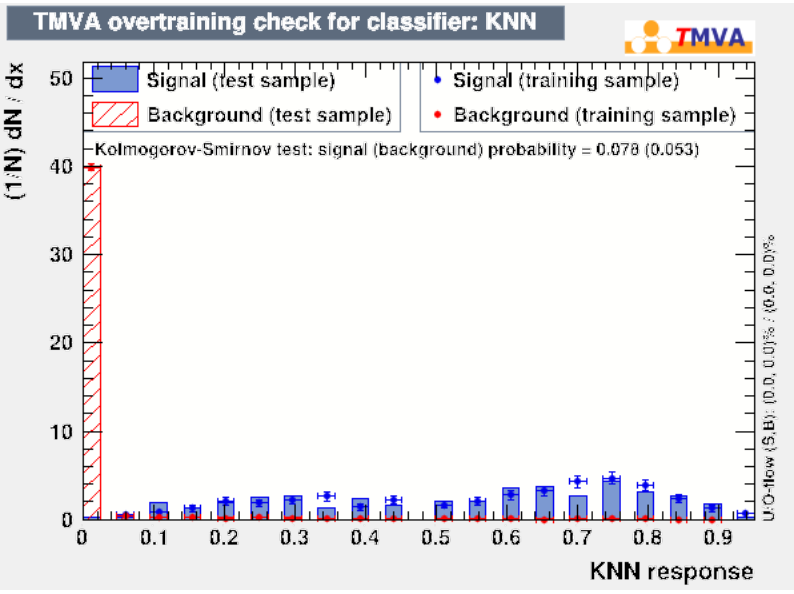


Figure 6: TMVA overtraining check for classifier: KNN

Conclusions

In the present work we have briefly explored MVA methods for signal to background discrimination. It was found that non linear methos, like BDT, kNN and MLPBNN offer the possibility to disentangle the signal better than most linear methods, and far better than possible with the application of a cut over a single variable. Specifically, from the input variable's distribtuions alone, it may seem obvious that the only useful variable was the lepton pair invariant mass. Nevertheless, the MultiVariate Analysis and Optimization takes advantage of the additional not automatically evident information to greatly improve the separation.

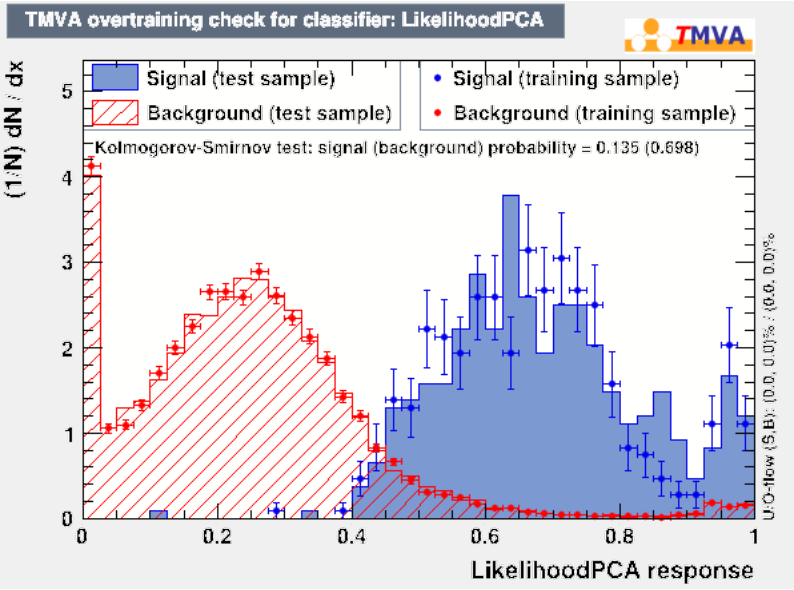


Figure 7: TMVA overtraining check for classifier: KNN

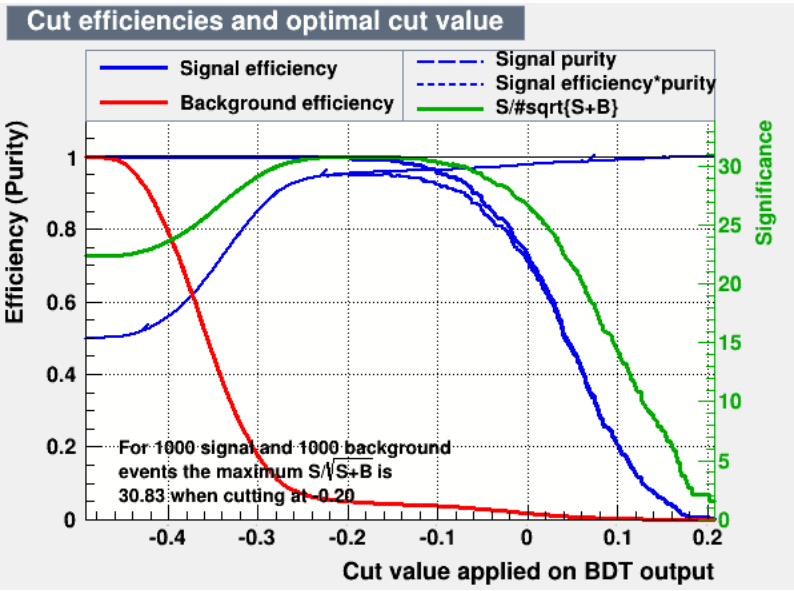


Figure 8: Cut optimization for classifier: BDT

On the other hand, we note that it is necessary to obtain sufficient amount of data to avoid overtraining. This applies specially for non linear methods, like BDT and and kNN. When dealing with interction channels that have an specially low cross section, like neutrino trident production, this may turn out to be problematic. However, we point out that the method that performed the best in our study was MLPBNN. It scores high in the background rejection/signal efficiency and yet doesn't fall in a significant overtraining (KS training vs test value $\sim$ 0,7)

MLPBNN is the better method multivariate

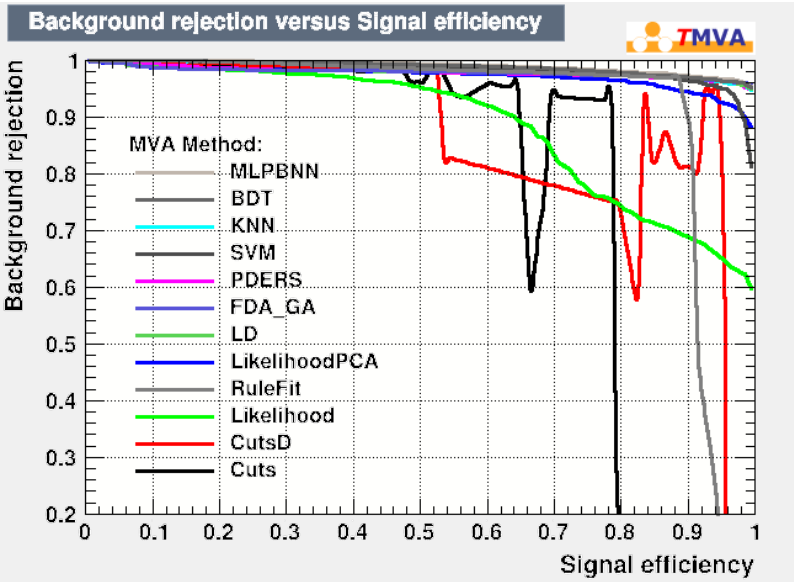


Figure 9: Background rejection versus signal efficiency

Appendix

2.1 Self Evaluation

The participation of the members was the same for all, in order to benefit and strengthen the skills of each member, including:

- Potentiation of prior knowledge, new skills and teamwork.

In the proposed dates for the stages of the project, we look at the weaknesses of the group, among which are:

- Data collection problems (errors sotfware, code, server, etc). - Availability, this is because of the course load and their respective "tasks". - The proposed scheme for the realization of the project.

Despite having had these weaknesses, we maintained constant communication through the "github" where we exchange the results for their respective discrimination.

2.2 Additional Plots

References

[1] HOECKER, A. et al. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT, p. 040, 2007.

[2] ADAMS, T. et al. Neutrino trident production from NuTeV. In: *High-energy physics. Proceedings, 29th International Conference, ICHEP'98, Vancouver, Canada, July 23-29, 1998. Vol. 1, 2.* [S.l.: s.n.], 1998.

[3] ADAMS, T. et al. Evidence for diffractive charm production in muon-neutrino Fe and anti-muon-neutrino Fe scattering at the Tevatron. *Phys. Rev.*, D61, p. 092001, 2000.

[4] BERGSMA, F. et al. Search for Coherent Muon Pair Production by Neutrinos and Anti-neutrinos. *Phys. Lett.*, B122, p. 185, 1983.

[5] GEIREGAT, D. et al. First observation of neutrino trident production. *Phys. Lett.*, B245, p. 271–275, 1990.

[6] MISHRA, S. R. et al. Neutrino tridents and W Z interference. *Phys. Rev. Lett.*, v. 66, p. 3117–3120, 1991.

[7] BELUSEVIC, R.; SMITH, J. W - Z Interference in Neutrino - Nucleus Scattering. *Phys. Rev.*, D37, p. 2419, 1988.

[8] ANTCHEVA, I. et al. Root — a c++ framework for petabyte data storage, statistical analysis and visualization. *Computer Physics Communications*, Elsevier BV, v. 180, n. 12, p. 2500–2512, Dec 2009. ISSN 0010-4655. Disponível em: <http://dx.doi.org/10.1016/j.cpc.2009.08.005>.

[9] ANDREOPOULOS, C. et al. The GENIE Neutrino Monte Carlo Generator: Physics and User Manual. 2015.