Particle Identification at DØ using Neural Networks

D. Chakraborty

Department of Physics and Astronomy, State University of New York at Stony Brook, Stony Brook, NY 11794-3800
(for the DØ collaboration)

Abstract. We have investigated the possibility of employing neural networks for identification of electrons and τ leptons with the DØ detector in the upcoming run of the Tevatron collider at Fermilab. Preliminary results based on Monte Carlo simulations indicate that for any acceptable level of signal efficiency, neural networks consistently outperform covariance matrices so far employed for the same purpose. Using a subset of variables used by a covariance matrix, a properly trained neural network offers 2 times better background rejection for taus, and 10 times for electrons, at 90% signal efficiency. Similar enhancements can be expected for identification of other objects (such as muons, *b* or *c* jets, quark vs gluon jets, neutrinos etc).

INTRODUCTION

Effective particle identification is a central challenge in Experimental High Energy Physics. Most of the interesting processes are buried in enormous backgrounds. "Smart" algorithms are needed to identify particles that feature in signal but not in background. Here we address the identification of electrons and τ leptons which appear in the final states of many important processes at the Tevatron collider experiments at Fermilab. During Run 1 (1992-96), the Tevatron delivered ~ 125 pb^{-1} of proton-antiproton collisions at a center-of-mass energy of 1.8 TeV to each of its two detector experiments, DØ and CDF. After an extensive upgrade to both the collider and the detectors, collisions will resume in the spring of 2001 (Run 2). The collider upgrade will result in a higher center-of-mass energy (2.0 TeV), and an integrated luminosity of $2-20$ fb^{-1} per experiment. The upgraded detectors will have superior momentum and position resolution.

DØ is a general-purpose detector (1). In Run 2, it will have a central cylindrical tracking volume consisting of silicon and scintillating fiber trackers immersed in a coaxial magnetic field of 2 Tesla where the trajectory of a charged particle can be mapped with a precision of a few microns, and the magnetic field will allow momentum measurement from the curvature of the track. Surrounding the tracker and the solenoidal magnet are preshower detectors and calorimeters providing an almost full solid angle coverage of the interaction point. The active medium of the calorimeter (preshower detector) is liquid argon (plastic scintillator) while the passive absorbers are made of depleted uranium or steel (lead). With the exception of muons and neutrinos, all particles leaving the tracking volume deposit all of their energy in the calorimeter. Momenta of the muons, which penetrate the calorimeter, are measured by tracking them before and after their passage through a thick layer of magnetized iron. The presence of any neutrino in an event can only be inferred from the imbalance in the transverse component of the total momentum. In Run 1, covariance matrices were employed to identify electrons and τs using several measurements obtained from different parts of the detector. Here we examine the merits of using artificial neural networks as an alternative to covariance matrices in the context of Run 2.

THE COVARIANCE MATRIX APPROACH

For a set of characteristic variables $\{x_1, x_2, \ldots x_n\}$, the $\{i,j\}$th element of the covariance matrix is given by

$$\sigma_{i,j} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle, \tag{1}$$

determined from a training sample of signal or background, either simulated or obtained from test-beam. For a given candidate with corresponding characteristics $\{X_1, X_2, \ldots X_n\}$, we define

$$\chi^2 = \sum_{i,j} X_i H_{i,j} X_j, \tag{2}$$

where H, also called "H-matrix", is the inverse of the covariance matrix:

$$H_{i,j} = (\sigma^{-1})_{i,j}. \tag{3}$$

For a suitably chosen set of characteristics, the χ^2 can be used as a multivariate discriminant to separate signal from background. If the matrix is trained on a signal sample, then, on average, χ^2s for signal will be smaller than those for background. This approach involves a simple linear transformation which is optimal if the variables follow a multivariate Gaussian distribution.

THE NEURAL NETWORK APPROACH

For the problems in our hands, however, the measured characteristics have asymmetric distributions and bear non-linear correlations. Neural networks, which can handle such features, are therefore expected to offer superior separation between signal and background. Specifically, we have tested the MLPfit neural network (2). We used a single layer of hidden variables or neurons. The jth neuron performs a linear combination of all the inputs,

$$h_j = c_j + \sum_i w_{i,j} x_i, \tag{4}$$

and delivers the output after a sigmoidal transformation

$$s(x) = (1 + e^{-x})^{-1}. \tag{5}$$

The output of the network is a linear combination of the outputs of the neurons:

$$y = k + \sum_j p_j s(h_j). \tag{6}$$

The network determines the constants $c_i, w_{i,j}, p_j$, and k, by fitting the functions to multidimensional distributions of input variables from samples of signal and background. We opted for a hybrid-linear Broyden, Fletcher, Goldfarb, Shanno (BFGS) method for minimizing the error function (3).

IDENTIFICATION OF ELECTRONS

We focus on identifying high-energy central electrons. Specifically, we require $|\eta| < 1.0$ and $E > 10$ GeV where $\eta \equiv -\ln(\tan\frac{\theta}{2})$ is the pseudorapidity with θ being the polar angle with respect to the beam axis, and E the energy of the electron. Our training samples consist of 2600 Monte Carlo single electrons, uniformly distributed in 10 GeV $< E <$ 100 GeV, for signal, and 1600 QCD dijet events with at least one jet hadronizing into a final state with π^0 fraction greater than 65% for background (hadronic jets with smaller π^0 fraction are not likely to mimick an electron). Similar samples were used for testing.

The energy deposited by an electron is contained in a short and narrow volume in the calorimeter. The DØ calorimeter is finely segmented in transverse as well as longitudinal direction. The 4 innermost layers of

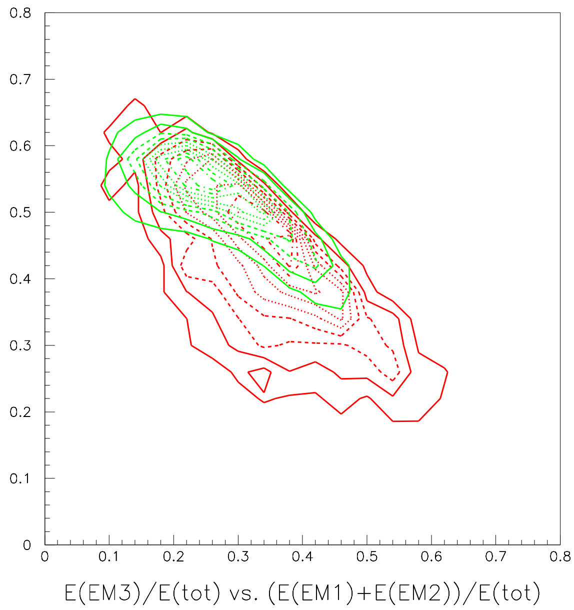

FIGURE 1. Correlation between energy fractions in EM layer 3 and the sum of layers 1 and 2 for signal (light contours) and background (dark contours).

the calorimeter, EM1, …EM4, and the central preshower detector (CPS) contain most of an electron's energy, while the shower widths in the azimuthal angle (ϕ) and pseudorapidity (η) give a measure of the lateral spread. The following 7 variables are given to the network: $E_{\mathrm{EM1+EM2}}/E$, E_{EM3}/E, E_{EM4}/E, E_{CPS}/E, $E/100$, σ_ϕ, and σ_η. Figure 1 shows the non-linear correlation between two of these variables.

After some optimization, we adopted a network with 16 neurons in the hidden layer and trained it over 400 epochs. The covariance matrix, or H-matrix, had a much more detailed picture consisting of 41 variables. Figure 2 shows the comparison between the H-matrix and the neural network results. We find that for any acceptable level of signal efficiency, the neural network affords us a level of background suppression 5 to 10 times better than that of the H-matrix.

IDENTIFICATION OF τ LEPTONS

The preselection of τ lepton candidates are similar to those of electrons: $|\eta| < 1.0$ and $E_T > 10$ GeV where E_T is the transverse component of the momentum. Our training samples consist of about 1150 hadronically decaying τs chosen from 2000 Monte Carlo $Z \to \tau\tau$ events based on the above criteria for signal, and about the same number of QCD jets for background. Similar samples are used for testing.

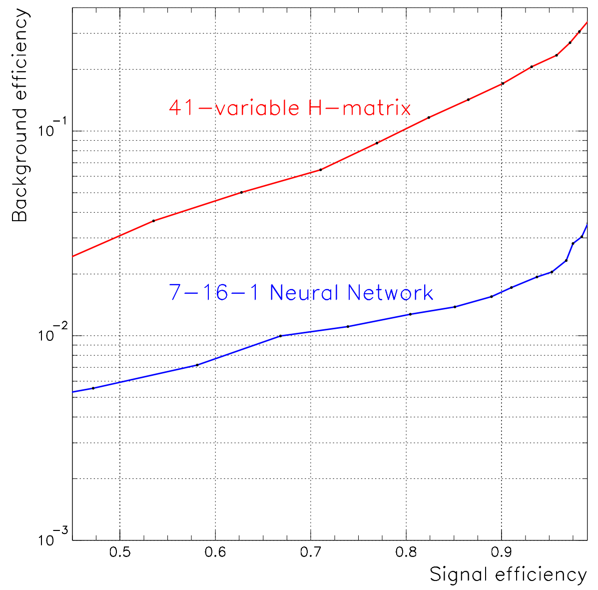

FIGURE 2. Fraction of background retained as a function of electron signal efficiency: for H-matrix (upper curve) and MLPfit neural network (lower curve).

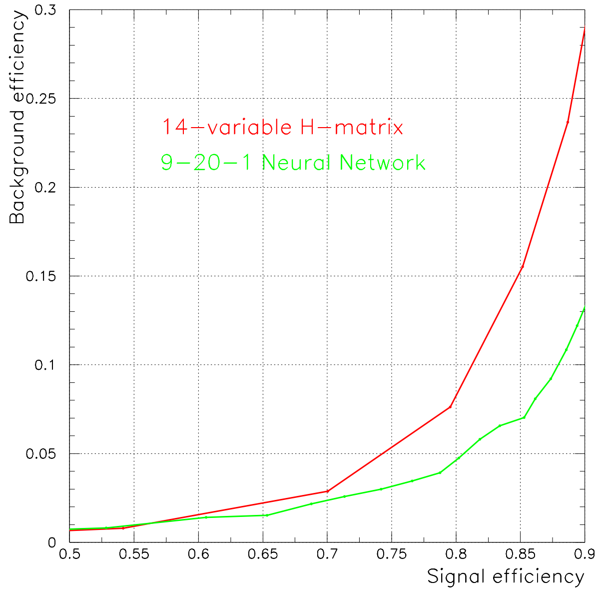

FIGURE 3. Fraction of background retained as a function of τ signal efficiency: for H-matrix (upper curve) and MLPfit neural network (lower curve).

Typically, the energy profile of a τ is narrower than that of a QCD jet although not as narrow as that of an electron. Also, on average, a τ has fewer associated tracks than a QCD jet. The following 9 variables are given to the tau neural network: $E_{\rm EM3}/E$, $E_{\rm EM4}/E$, $E_{\rm FH}/E$, RMS(cal), $E_{\rm hot2}/E$, $\log_{10}E-1$, $\log_{10}(N_{0.4}+1)$, $(N_{0.4}-N_{0.2})/(N_{0.4}+1)$, and RMS(trk), where FH refers to the first calorimeter layer after the four EM layers, and N_R is the number of tracks within a cone of radius R centered on the line joining the primary interaction vertex and the calorimeter cluster centroid. RMS(cal) is the root-mean-squared width of the calorimeter cluster in the $[\eta,\phi]$ space, while RMS(trk) is the momentum-weighted root-mean-squared width determined from the associated tracks with respect to the calorimeter cluster centroid. $E_{\rm hot2}/E$ is the fraction of the total energy in the two most energetic towers in the calorimeter cluster. This variable is the most powerful discriminant for τs. We adopted a network with 20 neurons in the hidden layer and trained the network over 500 epochs. The covariance matrix, or H-matrix, had 14 variables including all of the above except RMS(cal). Instead of RMS(cal), it had two other variables to quantify the lateral spread of the cluster. Figure 3 shows the comparison between the performances of the H-matrix and the neural network. For large signal efficiencies, the neural network admits only half as much background as the H-matrix, while performing comparably for smaller efficiencies.

CONCLUSION

Our studies indicate that the neural network consistently offers better background rejection for identification of electrons and τ leptons than the covariance matrix method. Preliminary studies suggest that further improvements can be expected from careful optimization. The findings also make a strong case for adopting neural networks for identification of other objects, such as muons, b and c jets, neutrinos, and separation of quark- and gluon-initiated jets.

REFERENCES

1. DØ Collaboration, S. Abachi et al. Nucl. Instr. and Methods A**338**, 185 (1994).
2. `http://www.cern.ch/~schwind/MLPfit.html`
3. P.E. Gill, W. Murray, M.H. Wright, "Practical Optimization" Academic Press, Inc., London (1981).