Analysing Particle Jets with Artificial Neural Networks

K.-H. Becks, J. Dahm and F. Seidel

Physics Department, University of Wuppertal
Wuppertal, Germany

Abstract. Elementary particle physics includes a number of feature recognition problems for which artificial neural networks can be used. We used a feed-forward neural network to separate particle jets originating from b-quarks from other jets. Some aspects such as pruning and overfitting have been studied. Furthermore, the influence of modifications in architecture and input space have been examined. In addition we discuss how self-organizing networks can be applied to high energy physics problems.

1 Introduction

During the last years there has been an increase of interest for "brain style computing" in terms of artificial neural networks (NN). The reason is the power NNs have shown for a wide variety of real-world feature recognition problems. The attractive features are adaptiveness, robustness, inherent parallelism, etc. In particle physics there are many challenging feature recognition problems such as quark flavour tagging\(^1\). Furthermore, with increasing luminosity and energy of new particle accelerators efficient extraction procedures will become more and more important. The standard procedure for extracting information from experimental data is performed by various cuts in parameter space. One wants to get the optimal choice of cuts which separate the different classes from each other. This is really what NN aim to do.

First a feed-forward NN with a backpropagation learning algorithm will be discussed. This architecture is used for classification of data where the features to be recognized are known beforehand. A feed-forward NN is an example of supervised learning. Secondly, we shortly explain how self-organizing networks, which form their own classification of the input patterns as the features to be recognized are unknown, could be used to detect new physical properties. Details about both kinds of networks can be found in [1].

The feed-forward networks are used in the context of high energy physics data analysis, e.g. for the quark flavour tagging problem. To understand this task better we shortly discuss the physics problem: Quark flavour tagging means the reconstruction of the initial quarks, the basic building blocks of nature, created in a particle collision process. The reconstruction process has to be performed from the observed particles in a detector system. In the relatively simple case of an electron-positron collision process...
annihilation experiment, a quark/antiquark pair (of the same flavour) might be created. The different quark flavours are up, down, strange, charm, beauty, and the (to now undetected) top quarks. The process of event reconstruction is complicated by the so-called fragmentation process of quarks into hadrons, by decays of short living hadrons into stable particles, and by clustering of single hadrons to particle jets. These complications make a direct way back from observed hadrons to the initial quarks impossible.

2 Quark Flavour Tagging

One main topic of interest for this study is "heavy flavour physics". For precise measurements of b-quark parameters one has to be able to distinguish whether a hadron jet originates from a b-quark or a light quark (u, d, s, c).

To solve this problem, we used two feed-forward network architectures with the standard backpropagation algorithm [2]. A layered feed-forward network with 40 input nodes, 20 hidden nodes, and 1 output node, as well as a network with two hidden layers, 40 input nodes, 20 hidden nodes in the first layer, 10 nodes in the second hidden layer, and 1 output node. The data used in our analysis were generated with a physics generator plus a full detector simulation. The train set consisted of 10000 event patterns from which the following simple input variables have been used:

- the absolute values of the momenta of the ten fastest particles in the two most energetic jets, called jet 1 and jet 2;
- the transversal components of momenta of the ten leading particles in jet 1 and jet 2 with respect to the jet axes.

From a physical point of view the two jets are almost independent. Taken this into consideration we have split the input layer in two halves, each receiving the variables of one jet. The aim was to force the net to deal with separated jets. It turned out that such a partially connected network provides a better performance than a fully connected one. To allow a comparison of the network performance we defined:

\[
\text{b-efficiency} = \frac{\text{number of correctly classified b}}{\text{number of all b events}}
\]

\[
\text{b-purity} = \frac{\text{number of correctly classified b}}{\text{number of patterns classified as b}}
\]

To achieve a further improvement of performance, we have included event shape features to the input patterns for the two hidden layer network. These features are the relative orientations of the three most energetic jets to each other.

Fig. 1 shows the result for such partially connected networks: one hidden layer (open squares), two hidden layers without relative orientations (black triangles) and with orientation (open circles).