Neural network technique for point source search : comparison with sequential cut method

- Introduction
- Exploitation of the PAW Neural Network
- Performances of the Neural Network
- Concluding remarks and prospect

Introduction

- Rejection of background and selection of signal in a sample of data: use of a neural network (multi-layer perceptron available in PAW)

- Comparison Neural Net ↔ optimized classical sequential method ("pedestrian method")

- Pedestrian method : cuts on 8 analysis variables

  - Zenith(2)
  - Ndirb(2)
  - Ndirc(6)
  - Ldirb(6)
  - Smootallrl(6)
  - Ndira(2)
  - Smootallphit
  - Zenith(6)

Results:

$\varepsilon_s = 9\ \%$

$\varepsilon_b = 0.005\ \%$

$R = 1650$

Exploitation of the PAW neural network:

- same analysis variables are used as input to the neural network (goal = comparison)

- determination of the performances by computing

- signal efficiencies (ε_s)
- background efficiencies (ε_b)
- the ratio $R = \varepsilon_s / \varepsilon_b$
- the error function E at the end of training

for

- different configurations of the net
- different learning methods available

→ to be optimized!

Configuration parameters

- Number of input neurons: 8
- Number of hidden layer neurons: variable
- Number of output neurons: 1

Training of the neural network has been made using equally populated samples of background and signal events (4000 each)

The range of variation of the different analysis variables has been normalized to (0 , 1)

The number of learning steps is 1000.

Remarks:

- Guidelines of the authors

- "In principle, one hidden layer is sufficient …"
- "There are no rules to choose the number of hidden neurons… empirical way… by following the evolution of the error with the numbers of *learning steps*…"
- "*To avoid overfitting the learning samples should be large enough*… There seems to be no strict rule on the ratio Number of *events*/Number of weights, which should be between 10 and 100"
- "It is better to start with a small number of neurons: learning is faster, often enough, avoids overfitting problems"

- Error function

The error function E which has to be minimized in order to find the best values of the neuronic weigths insuring efficient separation between background and signal is constructed as follows:

$$E = \frac{1}{2}\sum_p w_p\left(o_p - t_p\right)^2$$

where p runs over all the events contained in the training samples, o_p is the output value for event p, w_p is a weight taking into account the relative population of the training samples and t_p is the expected output value (0 if the event p is background; 1 if signal). As the training samples are equally populated, w_p is set equal to 1 for all events.

Performances of the Neural Network

- Evaluation of relative performances : cut set on NN output such that efficiencies comparable to pedestrian method

Performances as a function of learning method

(see table 1and fig. 1 a,b,c)

Fig. 1 shows:

- The distribution of events as a function of neural output value for signal, background and data (a)
- The evolution of the error function as a function of number of learning step (b)
- The efficiencies ε_i, the efficiency ratio $R = \frac{\varepsilon_s}{\varepsilon_b}$, the difference $(\varepsilon_s - \varepsilon_b)$ and the product $\Pi = R \times (\varepsilon_s - \varepsilon_b)$ for signal, background and data selection above cut on the neural output value (c)

→ Learning method n° 4 happens to be the best

Performances as a function of number of hidden layer neurons

(see table 2)

Although learning method n° 4 happened to be the best, learning method n° 1, slightly worse, has been adopted for practical reasons for these tests

Performances as a function of the ratio $\frac{w_s}{w_b}$ for the training of the neural network

(see table 3)

Remarks :

- Surprising : neural network output range differs generally from the expected (0, 1) range!

- Learning method n° 4 seems to be the best. Learning method n° 1 and n°3 would be acceptable.

- When the number of input neurons is 8, hidden layers with 10 or 20 neurons lead to similar performances

- Within the here investigated limits of variation of the configuration parameters → NN technique happens to be (much) less efficient, as far as the ratio R is concerned, than the sequential approach.

- It is to be expected that the physical distributions relative to the events selected either by the sequential method or by the neural network method would be different. The zenith or declination angle distributions and the sky map are of particular interest: they appear to be quite different:

    - Fig. 2 shows the distribution as a function of Zenith(6) of the AMANDA data events selected by the sequential cut method (black) and by the neural network technique
    - Fig. 3 shows the sky map distributions with neural network events (a) and with pedestrian events (b).

<u>Concluding remarks and prospect</u>

- NN technique less efficient than sequential cuts (until now)

- Resulting distributions of the selected events as a function of the zenith angle are quite different (while both analysis made use of the same analysis variables)

- Use of additional analysis variables is to be investigated !