# FIT5186 Intelligent Systems    Assignment

# Methods for multivariate particle classification

## Contents



| | | |
|---|---|---|
| **Student ID:** | **2469****** | **2469****** |
| **Student Name:** | **Shao Qi** | **Zhang Hongliang** |

## 1. Abstract

We present results from a case study comparing different multivariate to demonstrate the techniques and methodologies by using NeuroShell2 application. The input is a set of data which is generated to simulate registration of high energy gamma (γ) particles in a ground-based atmospheric Cherenkov gamma telescope using the imaging technique. Such a set of data belongs to two classes, originating either from incident γ-ray or caused by hadronic showers. Such a case study is constructed from the initial raw data sets, showing the required preprocessing and decision relating to architecture and parameters. Several methods give the final results and the results are presented and evaluated. The classification accuracy is evaluated by using Multilayer Feedforward Neural Network (MFNN) with 10-30-1 and 10-30-2 architecture.

Keyword: **Multivariate**, **NeuroShell2, Classification, MFNN**.

## 2. Introduction

Astronomy, astrophysics, and particle physics all have improved a big step as observational sciences in recent years. Much of these improvements are due to the development of the detector technology, coupled with a parallel development of analysis methods (Bock et al., 2004) . The γ -ray is widely used in astronomy and it is observed to research interstellar substance distribution, celestial body state and high-energy interaction on celestial body. A question on γ-ray is that classifying and isolating hadrons from γ-ray to improve signal-to-noise ratio of γ-ray observation. The spatial and temporal properties of Cerenkov photons contain valuable information. A ground-based atmospheric Cherenkov gamma telescope can capture images to observe γ-ray. Effectively excludes hadrons in γ-ray is a technique to improve the signal to noise ratio in the data.

Cherenkov gamma telescope is one of the detector technologies and observes high energy γ-rays, taking advantage of the radiation emitted by charged particles produced inside the electromagnetic showers initiated by the gammas (Lessard, Buckley, Connaughton, & Le Bohec, 2001) . This Cherenkov radiation leaks through the atmosphere and gets recorded in the detector, allowing reconstruction of the shower parameters. The following sections describe very briefly some problems arising in the analysis of their data, and the general characteristics of signal and background events in such devices.

## 3. Data Sets

For our case study, we used data sets generated by a Monte Carlo (MC) program, which is described by Heck, Knapp, Capdevielle, Schatz, and Thouw (1998). The data are MC generated to simulate registration of high energy gamma particles in a ground-based atmospheric Cherenkov gamma telescope using the imaging technique. Cherenkov gamma telescope observes high energy γ-rays (Moriarty & Samuelson, 2000), taking advantage of the radiation emitted by charged particles produced inside the electromagnetic showers initiated by the gammas, and developing in the atmosphere.

In the work presented here we perform a general study, applying neural network classification method to events described by 10 chosen image parameters, without claiming the parameters to be optimal for these events. I find these data and the project from the website http://archive.ics.uci.edu/ml/datasets/MAGIC+Gamma+Telescope. The data consist of two classes: gammas (signal) and hadrons (background). Events were generated at shower energies from 10 GeV up to about 30 TeV; and for zenith angles from zero to 20. Each event is characterized by the following 11 attributes (include the class):

| | | | |
|---|---|---|---|
| 1 | flength: | continuous | major half axis of ellipse [mm], |
| 2 | fwidth: | continuous | minor half axis of ellipse [mm], |
| 3 | fsize: | continuous | 10-log of sum of content of all pixels [photon count], |
| 4 | fconc: | continuous | ratio of sum of two highest pixels over size [ratio], |
| 5 | fconc1: | continuous | ratio of brightest pixel over size [ratio], |
| 6 | fpdist: | continuous | distance from brightest pixel to center, along major axis [mm], |
| 7 | fm3long: | continuous | 3rd root of third moment along major axis [mm], |
| 8 | fm3trans: | continuous | 3rd root of third moment along minor axis [mm], |
| 9 | falpha: | continuous | angle of major axis with vector to origin [deg], and |
| 10 | fdist: | continuous | distance from origin to center of ellipse [mm]. |
| 11 | class: | g, h | # gamma (signal), hadron (background) |

The original data file contains the detail data, together with the observed classification of each signal event as a gamma event or hadrons event. There are 1000 rows in the data file, comprising 700 known gamma events and 300 known hadrons events. This data has been preprocessed to replace categorical data with numerical data. Neural networks need to receive numerical data as inputs, and the data needs to contain a sense of scale and order. By this is meant that a neural network will treat inputs of 1 or 2 as being similar to each other but different from say 9 or 10.

The original data contain 10 columns or parameters: 10 numerical continuous data columns (eg. flength or fwidth) and 1 categorical data column (class). The categorical data needs to be converted to numerical. The known classification of each event in the data set as ‘g’ or ‘h’ can be coded as either: class 1 or 2 (using a single column) or using two columns to encode gamma (1 0) or hadron (0 1). An example showing the transformation of the first 5 rows of the data after preprocessing is showing in the table1.

**Table 1**: SAMPLE AFTER PRE-PROCESSING OF DATA

| fLength | fWidth | fSize | fConc | fConc1 | fAsym | fM3Long | fM3Trans | fAlpha | fDist | class | gamma | hadron |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 28.7967 | 16.0021 | 2.6449 | 0.3918 | 0.1982 | 27.7004 | 22.011 | -8.2027 | 40.092 | 81.8828 | 1 | 1 | 0 |
| 31.6036 | 11.7235 | 2.5185 | 0.5303 | 0.3773 | 26.2722 | 23.8238 | -9.9574 | 6.3609 | 205.261 | 1 | 1 | 0 |
| 162.052 | 136.031 | 4.0612 | 0.0374 | 0.0187 | 116.741 | -64.858 | -45.216 | 76.96 | 256.788 | 1 | 1 | 0 |
| 23.8172 | 9.5728 | 2.3385 | 0.6147 | 0.3922 | 27.2107 | -6.4633 | -7.1513 | 10.449 | 116.737 | 1 | 1 | 0 |
| 75.1362 | 30.9205 | 3.1611 | 0.3168 | 0.1832 | -5.5277 | 28.5525 | 21.8393 | 4.648 | 356.462 | 1 | 1 | 0 |

## 4. Training Issues and Results

NN has been presented frequently in the past; ANN-s resemble the tree based methods in that they define simultaneous selections in variables, but instead of the original variables they work in locally linearly transformed data, and the transformation itself is part of the optimization (learning) process. We now have a choice of trying to classify signal event according to their class number (1 for gamma or 2 for hadron) using a single output neuron, or using two output neurons. In this research, the study uses NeuroShell2 as a tool to train and learn to obtain a goal: classify gammas and hadrons. The following four experiments are used to compare and analyze which architecture is the best.

### *Experiment 1*

In the first experiment, we choose all 10 variables from the preprocessed data file as inputs, and construct a MFNN architecture consisting of 10 inputs, 30 hidden neurons and 2 output neurons. 20% of the data is randomly extracted to form the test set. The learning rate and momentum rate are fixed at c=0.1 and α=0.1 while the initial weights are small random numbers around 0.3. The order in which input patterns are presented t to the MFNN is random, and the network performance is measured on the test set every 200 epochs. If the test error has not improved within 20000 epochs, then the training is terminated.

We decide that if the output of the gamma is greater than the out of the hadrons, then the event belongs to the gamma event, and vice versa. Using this method, and evaluating the performance across the entire data set, the neural network is able to correctly classify 672 of the 700 known gamma events and 102 of the 300 hadrons events. The result are summarized in Table2 showing the accuracy levels of the decisions.

**Table2**: Classification accuracy using MFNN with 10-30-2 architecture

| | classified gamma | classified hadron | row accuracy |
|---|---|---|---|
| actually gamma | 672 | 28 | 96.0% |
| actually hadron | 198 | 102 | 34.0% |
| column accuracy | 77.2% | 78.5% | |

Thus the trained neural network can be expected to accurately classify gamma events 96.0%, but only recognize a hadrons event 34.0%. Obviously it is more difficult for the neural network to learn to recognize the characteristics of hadrons event. However, 78.5% of the classified hadrons are actually hadrons and the result is very good. In the second experiment, we should improve the percentage of recognizing hadrons.

## *Experiment 2:*

Looking at the contribution each variable is making to the decision making process (NeuroShell2 enables this to be seen in graphical form) reveals that some parameters are extremely significant to the output of the network, while other have little effect. Parameters like flength, fwidth, fm3long, fm3Trans and falpha are found to have the greatest effect. In this experiment, we will use only these 5 attributes for each input pattern rather than the set of all 10 attributes. All other parameters are set as before. The result for this experiment is shown in Table3.

**Table 3**: Classification accuracy using MFNN with 5-30-2 architectureTable3

| | classified gamma | classified hadron | row accuracy |
|---|---|---|---|
| actually gamma | 676 | 24 | 96.6% |
| actually hadron | 162 | 138 | 46.0% |
| column accuracy | 80.7% | 85.2% | |

The effect of removing unnecessary variables from the network has been to slightly improve the percentage of hadrons that are detected (46.0% rather than 34.0% as before). In the above two experiment, the number of hidden neurons are both 30. What happens if we change the number of the hidden neurons?

## *Experiment 3*

Further experimentation with learning rates and momentum rates does not significantly improve the results. However, different hidden neurons may have some effect. So in the third experiment, we have varied the number of hidden neurons from 30 to 20 and 40 neurons, still with the reduced 5 inputs. The results are shown in Table 4 and Table 5 for 20 and 40 hidden neurons respectively.

**Table 4**: Classification accuracy using MFNN with 5-20-2 architecture

| | classified gamma | classified hadron | row accuracy |
|---|---|---|---|
| actually gamma | 675 | 25 | 96.4% |
| actually hadron | 173 | 127 | 42.3% |
| column accuracy | 79.6% | 83.6% | |

**Table 5**: Classification accuracy using MFNN with 5-40-2 architecture

| | classified gamma | classified hadron | row accuracy |
|---|---|---|---|
| actually gamma | 675 | 25 | 96.4% |
| actually hadron | 171 | 129 | 43.0% |
| column accuracy | 79.8% | 83.8% | |

Looking at the above two tables, we find that neither architecture appears to improve the performance of the network, nor so we will revert to the original 30 hidden neurons for following experiments.

## *Experiment 4*

At this stage, it is natural to wonder if any changes in architecture, inputs or parameters will improve the classification accuracies. In the above experiments, we choose two neuron outputs. So if we choose only one neuron output, what the classification accuracy will be. In this study, the gamma and hadrons are coded as 1 and 2 respectively. For the single output neuron architecture, we decide that the event likely belongs to gamma if the output of the network is less than 1.5 and hadrons if the output is greater than 1.5 (as we have chosen 1 and 2 to represent each signal). Besides, we also use a different threshold (not 1.5) to check what will happen. Here we choose other two thresholds: 1.3, 1.7. We will get three different results using different thresholds and the results are summarized in following three tables.

**Table 6**: Classification accuracy using MFNN with 5-30-1 architecture and decision threshold of 1.5

| | classified gamma | classified hadron | row accuracy |
|---|---|---|---|
| actually gamma | 678 | 22 | 96.9% |
| actually hadron | 165 | 135 | 45.0% |
| column accuracy | 80.4% | 86.0% | |

**Table 7**: Classification accuracy using MFNN with 5-30-1 architecture and decision threshold of 1.3

| | classified gamma | classified hadron | row accuracy |
|---|---|---|---|
| actually gamma | 602 | 98 | 86.0% |
| actually hadron | 223 | 77 | 25.7% |
| column accuracy | 73.0% | 44.0% | |

**Table 8**: Classification accuracy using MFNN with 5-30-1 architecture and decision threshold of 1.7

| | classified gamma | classified hadron | row accuracy |
|---|---|---|---|
| actually gamma | 692 | 8 | 98.9% |
| actually hadron | 170 | 130 | 43.3% |
| column accuracy | 80.3% | 94.2% | |

Analyzing from the above three tables, the smaller the threshold is, the more hadrons we can detect (the number of hadrons, table 7>6>8). From the above three tables, we find threshold of 1.7 has a more ability to recognize whether it is hadrons or gamma. However, for different problems, the specific threshold may be different and we should try different threshold to improve the classification accuracy.

## 6. Limitations

For technical reasons, the number of h events is underestimated. In the real data, the h class represents the majority of the events. The simple classification accuracy is not meaningful for this data, since classifying a background event as signal is worse than classifying a signal event as background. No classification method itself can substitute improvements in pre-processing of the image. For such a matter, invent new, independent parameters containing more information is necessary for future study. They may be derived from the image, but could also be derived from new, independent observations. A more efficient multivariate approach can be employed after the limitations are solved (Chilingaryan, Gevorgyan, Vardanyan, Jones, & Szabo, 2002).

## 7. Conclusion

Some of observational effects are taken care of by the transformation from pixel contents to parameters according to the Cherenkov telescopes. None of these distortions of the parameter space have been the subject of our study. We observe from the above results that the results from different architectures are close to each other. We also have use seven architectures to find which the best is. The methodologies under study are shown to be superior to direct cuts in parameters. So we use 5 parameters for comparison. After comparing, we can conclude that the 5-30-2 architecture and the 5-30-1 architecture with the threshold 1.7 are both very good. Here we choose 5-30-2 architecture as our best experiment. The conclusions that can be drawn from the results are valid for our input data. The classification operations all use MFNN architecture and results from each calculation are automatically generated by NeuroShell2.

## References

Bock RK., Chilingarian A., Gaug M., Hakl F., Hengstebeck Th., Jiřina M., & Towers S. (2004). Methods for multidimensional event classification: a case study using images from a Cherenkov gamma-ray telescope. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 516*(2), 511-528.

Chilingaryan A., Gevorgyan N., Vardanyan A., Jones D., & Szabo A. (2002). Multivariate approach for selecting sets of differentially expressed genes. *Mathematical Biosciences, 176*(1), 59-69.

Heck D., Knapp J., Capdevielle JN., Schatz G., & Thouw T. (1998). *CORSIKA: A Monte Carlo code to simulate extensive air showers* (Vol. 6019): FZKA.

Lessard RW., Buckley JH., Connaughton V., & Le Bohec S. (2001). A new analysis method for reconstructing the arrival direction of TeV gamma rays using a single imaging atmospheric Cherenkov telescope. *Astroparticle Physics, 15*(1), 1-18.

Moriarty P., & Samuelson FW. (2000). *Kernel analysis in TeV gamma-ray selection.* Paper presented at the AIP Conference Proceedings.