Publishers' page

Publishers' page

Publishers' page

Publishers' page

to World Scientific Publishing Company

MULTIVARIATE METHODS IN HIGH ENERGY PHYSICS
The Neural Network Revolution

By

Pushpalatha C. Bhat and Harrison B. Prosper

Preface

High energy physicists record data from billions of collisions of sub-nuclear particles in the hope of making major advances in our understanding of the Universe. It is possible that discoveries will come easily. Physicists have been lucky. It is more likely, however, that finding the signals of new physics could be a veritable case of "finding a needle in a hay-stack." If so, we believe the use of optimal methods of data analysis will be crucial to find these needles and, if we are especially fortunate, the odd jewel. Since high energy physics processes are generally characterized by many variables, optimal methods are necessarily multivariate.

Our interest in multivariate methods grew from our involvement in the discovery of the top quark in 1995, as members of the DØ collaboration. As early as 1990, it became clear to us that multivariate methods were potentially superior to those in widespread use in high energy physics and therefore warranted the effort required to understand how to use them effectively. The revolution in computational hardware and the development of elegant algorithms such as artificial neural networks made the work richly rewarding and fascinating. Also, the success of these methods in the hands of pioneering researchers around the world has convinced us that they will be the methods of choice in future analyses.

There is no shortage of monographs on the subject of multivariate analysis. Unfortunately, none is targeted specifically to topics of relevance to contemporary scientific analyses. There is quite a bit of information about these methods in the scientific literature. The information, however, is rather scattered, which makes absorbing it a daunting, and often confusing, proposition for a novice or a busy researcher. Our goal in writing this book is to draw together this dispersed wisdom as well as to provide a unified, coherent and practical exposition of multivariate methods as it is (and will be) practiced in experimental high energy physics. We hope that this book will serve as a reference, not only for seasoned particle physicists, but also for graduate students, postdoctoral fellows and researchers in other fields who need a clear but rapid introduction to the subject.

We have tried to make this book as self-contained as possible. The book has an implicit division into two parts. Chapters 1, 2 and 3 constitute the first part. These chapters contain an exposition of the main concepts of multivariate methods that are most relevant to practicing scientists. Our treatment of the subject is not exhaustive - to have attempted to make it so would have defeated our purpose. But it is, we believe, quite sufficient to cover most topics of interest in contemporary high energy physics research, as well as research in other scientific fields.

The first chapter provides the context for the kinds of applications to which multivariate methods have been successfully applied. Since high energy physics is our focus, we describe these methods using as a backdrop analysis problems of the sort that occur in that field. The probability theory needed to understand the methods described later, is also introduced in this chapter.

The second and third chapters are more technical. Here we provide a systematic exposition of various multivariate methods, from grid searches to neural networks. It is our experience that these methods are all too often presented as having intrinsically different purposes. However, in our view, they represent different approximations to the same ideal. In the remaining chapters our focus shifts to applications. We use real examples drawn from experimental high energy physics. Since our aim is to illuminate the common underlying principles behind these examples, we have, on occasions, taken the liberty to present them differently from the way they appear in the literature. With a unified perspective that we have tried to emphasize, of the various methods, one can more sensibly appreciate their differences, both their advantages and shortcomings and avoid coming to conclusions that are unwarranted. We have tried very hard to be as precise as we can without cluttering the text with, what most physicists would regard as, mathematical niceties. This is a practical book that aims for conceptual clarity; it is not a mathematical treatise. It is written to be enjoyed and to be used.

Contents

Chapter 1
Introduction

1.1 Experimental High Energy Physics

The twentieth century has seen monumental discoveries in science. *Modern Physics* which dealt with the atom at the beginning of the century has marched through the nuclear core of the atom and the innards of the nuclear constituents, the protons and neutrons (together call *nucleons*, evolving by the end of the century as *Elementary Particle Physics*, whose goal is to discover the ultimate constituents of all matter in the universe and their interactions. In essence, the aim is to acquire a comprehensive knowledge of fundamental matter and forces in the universe and their workings. Higher and higher energy experiments are needed to delve into the deeper realms of matter and to probe ever smaller distance scales. Hence came the name *High Energy Physics*, the exploration at the energy frontier, which seems to transcend space, time and energy, to truly seek the most elementary basis for all the complexities of the universe.

Then a paragraph or two about the current status of particle physics

The standard model, an elegant and concise theory of elementary particles and their interactions.

1.1.1 *What We Do and How We Do It*

All things we know for sure about the world, whether it be about quarks or about the cosmos, comes from experimental information. In this section, we provide some details of various experimental aspects of High Energy

Physics.

1. Describe goals of HEP
2. General description of experiments

The primary goal of a particle detector in high energy physics experiments, is to measure the energy and momenta of particles produced in the collision and to identify them if possible. The mammoth and complex particle detectors used in modern experiments are in fact an assemblage of a variety of particle detectors that serve to achieve this complex goal. In symmetric colliders, where the particles in the two colliding beams are similar in mass, the detectors are normally built in three sections - one central and two forward. In the central region, different types of detectors are built in concentric cylindrical shells around the interaction region, with roughly uniform segmentation in pseudorapidity (η) and azimuth (ϕ). To provide maximal coverage for measurement of collision products, layers of similar detectors are built in the forward and backward regions.

A typical collider detector consists of three major sub systems: an inner tracking system for precision tracking of charged particles emanating from the collision event; a sampling calorimeter that measures the total energy of electromagnetic and hadronic particles and an outer shell comprising muon detection system. Muons are massive cousins of electrons that only leave a minimally ionizing trail in inner detectors. The inner tracking system normally would have a core of silicon strip detecors. Surrounding the silicon detectors are the coarser particle tracking detectors that use either optical fibers or drift chambers. The calorimeter is made up of an electromagnetic part with large radiation length, and a fraction of an interaction length, followed by a hadronic part. The outer muon system is usually a magnetic spectrometer with toroidal magnetic field and layers of drift chambers arranged as inner and outer shell.

1.1.2 *High Energy Physics Analysis: A Primer*

1. Give general description

At various stages of data processing and analysis, beginning with the need to make a decision as to whether an event seen by the detector is worth keeping or not to the classification of events and their statistical analysis, multivariate methods can be gainfully employed. We discuss such applications in Chapter xx. The aim of this section is to provide a quick tutorial on how a high energy physics analysis is done.

We begin with the scenario where the data is already processed or "reconstructed." The reconstructed data would have, for each event, all possible measurements of particles and many deduced event characteristics and quantities, such as the four vectors of jets of particles, imbalance in transverse energy (referred to as the missing transverse energy).

2. A real-life example (top mass, leptoquarks)

1.2 Fundamental Concepts of Pattern Recognition?

1.2.1 *Elements of Probability Theory*

1. definition of probability
2. product and sum rule
3. bayes theorem
4. probability densities
5. some distributions

1.2.2 *Classification*

1. geometrical perspective
2. dimensionality reduction
3. optimal discrimination (bayesian theory)

1.2.3 *Parameter Estimation*

1. likelihood methods
2. bayesian methods

Chapter 2
Multivariate Methods

1. Introduce a simple 2-D model that can be used to illustrate each of the following methods.

2.1 Grid Searches

1. simple cuts; refer back to analysis primer
2. uniform grid search
3. random grid search; mention possibility of rotation of input variables into uncorrelated variables.

2.2 Binary Decision Trees

2.3 Gaussian Classifiers

If the covariance matrices for the 2 classes are different, the optimum discriminant function would be quadratic.

Start with likelihood function:

$$f(x) = A\ exp(-\frac{1}{2}(x-\mu)^T M^{-1}(x-\mu))$$

Bayes discriminant for S & B classes:

$$R = \frac{P(s1x)}{P(b1x)} = \frac{P(x1s)}{P(x1b)} * \frac{P(s)}{P(b)}$$

Fisher's discriminant:

$$\begin{aligned} F \quad = \\ & \log R = \\ & \log P(x1s) - \log P(x1b) + \log(\frac{P(s)}{P(b)}) = \\ & \frac{1}{2}(x - \bar{x_b})^T H_b (x - \bar{x_b}) - \frac{1}{2}(x - \bar{x_s})^T H_s (x - \bar{x_s} + \frac{1}{2}\log\frac{|H_s|}{|H_b|} + \log(\frac{P(s)}{P(b)}) \end{aligned}$$

H = M^{-1} n

$$F = \frac{1}{2}(X_b^2 - X_s^2) + \frac{1}{2}\log\frac{|H_s|}{|H_b|} + \log(\frac{P(s)}{P(b)})$$

If

$$H_s = H_b$$

and $P(s) = P(b) then F represents Fisher's linear discriminant$

2.4 Likelihood Methods

2.5 Probability Density Methods

2.6 Neural Networks

Chapter 3
Neural Networks

3.1 History

1. Give milestones pictorially (dates, people, ideas)

For most of human history, very little was known about the detailed structure of the brain and its workings. It was in the late nineteenth century that the structure of the brain was beginning to be unravelled. The nervous system was then claimed to be a network - either a continuous, uninterrupted web of nerve fibres (reticularist hypothesis) or composed of a very large number of discrete, interconnected cellular units called *neurons* (neuronalist hypothesis). The neuronalist view was proven in 1888 by a Spanish doctor Santiago Ramon y Cajal using a technique invented by Camillo Golgi in 1880. (Ramon y Cajal showed the presence of tiny gaps between individual neurons.) Since then, a lot has been learned about the brain and the internal structure of its neural networks; particularly amazing revelations have come since the advent of the electron microscope in 194**. (and interception and signal processing capability).

The human central nervous system is an incredibly complex system of neural networks - massively parallel, highly interconnected. Each neuron receives input signals from other neurons at its many *dendrites*, processes the information within the cell, and sends its output through a single tubular fibre, the *axon* which branches at its end to provide signal to other neurons (or muscle fibres). The transmission of information (signal) occurs at these ends of the axon, the *synapses*. The signals are transmitted electrically within a neuron and chemically at the synapses. The response of each neuron is a non-linear function of its inputs. The exact response

Table 3.1 Neural vs. Traditional Computing.

Neural Computing	Traditional Computing
Parallel Processing	Serial Processing
Learns from examples and generalizes	Needs pre-defined rules
Information distributed anarchic system	CPU controlled, Autocratic
Fault-tolerant degrades gracefully	Sensitive to memory loss
Does not break-down the problem logically	Programmed with logical approach
Content addressable memory	Random access memory

also depends on the processing neuron's internal state and input connection strengths. The networks of these non-linear processing units, therefore, can give rise to extremely complex and useful behaviours. They are adaptive, learn from experience and perform cognitive tasks.

Artificial neural networks, either simulated algorithms in computer programs or built in hardware, are paradigms in "the image of the brain." They exploit the massively parallel local processing and distributed representation properties that exist in hte biological neural systems. They also have many highly desirable features for information processing systems. They are adaptive, learn from examples/experience; are fault tolerant and can use noisy or fuzzy information. Salient features of neural computing systems are contrasted with traditional computing systems in table

3.2 Feed-Forward Neural Networks

3.2.1 *Bayesian Connection*

1. Give proof for arbitrary target values (other than 0 and 1)

2. Discuss PDE using NN

3. Bayesian interpretation of learning

Consider two (2) classes with likelihoods $P_1(\rightarrow x), P_2(\rightarrow x)$ such that $\int P_1(\rightarrow x)dx = \int P_2(\rightarrow x)dx = 1$

Proportions (priors) of (2) classes: $\mathrm{x}_1, x_2 with x_1 + X_2 = 1$

For 2-class problem, use NN with 1 output Error function: E=$1_{\overline{2\sigma}}$

3.2.2 *Algorithms for Training*

0. Error functions (many, but only some give Bayesian interpretation)

1. Give general statement of problem: to find the global minimum of the *true* error function, using the *actual* error function.

$$O_i^p(x) = g\Sigma_j\omega_{i_j}g(\Sigma\omega_{j_k})$$

p: patterns;
i: ith output
j: jth hidden node/neuron;
k: kth input
ik^p : kth input variable for patttern p
Training the network involves adjusting the weights ω_{i_j}, ω_{j_k} such that a given pattern p with inputs x^p yields an output $O_i^p \to$ desired value t_i^p.
This is done by minimizing the mean square error function,

$$E = \frac{1}{2}\sum_p\sum_i(O_i^p - t_i^p)^2$$

Use back-propagation of errors; adjust/update weights using gradient descent method.

3.2.2.1 *Back-Propagation*

$$E = \frac{1}{2}\Sigma_p\Sigma_i(Y_i^p - t_i^p)^2 = \frac{1}{2}\Sigma_i\Sigma_p[g\{\Sigma_j\omega_{i_j}g(\Sigma_k\omega_{j_k}x_kk^p)\} - t^p]$$

$$h_j = g(\Sigma_k\omega_{j_k}x_k) = g(a_j)$$

$$Y_i = g(\Sigma_j\omega_{i_j}h_j) = g(a_i)$$

$\Delta\omega_{i_j} = -\eta\frac{\partial E}{\partial\omega_{i_j}}$

$\Delta\omega_{j_k} = -\eta\frac{\partial E}{\partial\omega_{j_k}}$

$\eta = learningstrength$

$\frac{\partial E}{\partial\omega_{i_j}} = \Sigma_i\frac{\partial E}{\partial Y_i} * \frac{\partial Y_i}{\partial a_i} * \frac{\partial a_i}{\partial h_j} * \frac{\partial h_j}{\partial a_j} * \frac{\partial a_j}{\partial\omega_{j_k}} = \delta_i g^1(a_i)h_j; \delta_i = y_i - t_i$

$$\frac{\partial E}{\partial \omega_{jk}} = \Sigma_i \frac{\partial E}{\partial Y_i} * \frac{\partial Y_i}{\partial a_i} * \frac{\partial a_i}{\partial h_j} * \frac{\partial h_j}{\partial a_j} * \frac{\partial a_j}{\partial \omega_{jk}}$$

$$= \Sigma_i \delta_i g^1(a_i) \omega_{i_j} g^1(a_j) x_k$$

$$= g^1(a_j) X_k \Sigma_i \omega_{i_j} \delta_i g^1(a_i) = g^1(a_j X_k \Sigma_i \omega_{i_j} \delta_i g^1(a_i) = g^1(a_j) X_k \Sigma_i \omega_{i_j} \Delta_i$$

3.2.3 *Network Heuristics*

1. Pre-processing input variables
2. Choosing network architecture.
3. Training, stopping and testing.
4. Generalization
5. Choosing network parameters
6. Choosing input variables (PCA)

3.2.4 *Network Committees and Trees*

3.2.5 *Network Software*

3.2.6 *Other Networks*

Chapter 4
Applications in High Energy Physics

4.1 Measurement of the Top Quark Mass

4.2 Analysis of Top to Multijets

4.3 Searches for the Higgs Boson

4.4 Searches for Leptoquarks

4.5 Decays of the Z boson to heavy quarks

Chapter 5

Other Applications in High Energy Physics

5.1 Particle Identification

5.1.1 *Electron Identification*

5.1.2 *Tau Identification*

5.1.3 *b-jet tagging*

5.1.4 *Quark-Gluon Jet Discrimination*

5.2 Pattern Recognition

5.3 Triggering

Chapter 6

Examples of Applications in Other Fields

6.1 Astrophysics

6.2 Medical Physics

6.3 Solid State Physics

6.4 Accelerator Controls

Chapter 7

7 Hardware Neural Networks

7.1 What is Available and How it can be Used

7.2 What would be Nice to Have

Chapter 8

The Future