

New Machine Learning Developments in ROOT

S. V. Gleyzer¹, L. Moneta², Omar A. Zapata³, add your name here⁴

¹ University of Florida

² CERN

³ University of Antioquia and Metropolitan Institute of Technology

⁴ add you university

E-mail: Sergei.Gleyzer@cern.ch, Lorenzo.Moneta@cern.ch, Omar.Zapata@cern.ch, add your email here

Abstract. ROOT provides advanced statistical methods needed by the LHC experiments to analyze their data [?]. These include machine learning tools for classification, regression and clustering. TMVA, a toolkit for multi-variate analysis in ROOT [?], provides these machine learning methods.

1. Introduction

We will present new developments in TMVA, including parallelisation, deep-learning neural networks, new features and additional interfaces to external machine learning packages. We will show the new modular design of the new version of TMVA, cross-validation and hyper parameter tuning capabilities, feature engineering and deep learning. We will further describe new parallelisation features including multi-threading, multi-processing, cluster parallelisation and present GPU support for intensive machine learning applications, such as deep learning.

2. TMVA Interfaces

Interfaces to External ML Tools

- RMVA interface to R
- PyMVA interface to scikit-learn
- KMVA interface to Keras
 - High-level interface to Theano,
 - TensorFlow deep-learning libraries

2.1. RMVA interface to R

TODO by omar

2.2. PyMVA interface to scikit-learn

TODO by omar

2.3. KMVA interface to Keras

TODO by Stefan

3. Deep learning

TODO by Simon

4. Cross Validation

TODO by Thomas

5. Hyper-parameter tuning

TODO by Thomas

6. New regression features

TODO by Andrew

7. Jupyter Integration

TODO by Attila

7.1. `jpconf.cls` package options

The `jpconf.cls` class file has two options ‘a4paper’ and ‘letterpaper’:

```
\documentclass[a4paper]{jpconf}
```

or

```
\documentclass[letterpaper]{jpconf}
```

Table 1. `jpconf.cls` class file options.

Option	Description
<code>a4paper</code>	Set the paper size and margins for A4 paper.
<code>letterpaper</code>	Set the paper size and margins for US letter paper.

The default paper size is A4 (i.e., the default option is `a4paper`) but this can be changed to Letter by using `\documentclass[letterpaper]{jpconf}`. It is essential that you do not put macros into the text which alter the page dimensions.

8. The title, authors, addresses and abstract

The code for setting the title page information is slightly different from the normal default in \LaTeX but please follow these instructions as carefully as possible so all articles within a conference have the same style to the title page. The title is set in bold unjustified type using the command `\title{#1}`, where `#1` is the title of the article. The first letter of the title should be capitalized with the rest in lower case. The next information required is the list of all authors’ names followed by the affiliations. For the authors’ names type `\author{#1}`, where `#1` is the list of all authors’ names. The style for the names is initials then surname, with a comma after all but the last two names, which are separated by ‘and’. Initials should *not* have full stops. First names may be used if desired. The command `\maketitle` is not required.

The addresses of the authors’ affiliations follow the list of authors. Each address should be set by using `\address{#1}` with the address as the single parameter in braces. If there is more than one address then a superscripted number, followed by a space, should come at the start

of each address. In this case each author should also have a superscripted number or numbers following their name to indicate which address is the appropriate one for them.

Please also provide e-mail addresses for any or all of the authors using an `\ead{#1}` command after the last address. `\ead{#1}` provides the text Email: so #1 is just the e-mail address or a list of emails.

The abstract follows the addresses and should give readers concise information about the content of the article and should not normally exceed 200 words. **All articles must include an abstract.** To indicate the start of the abstract type `\begin{abstract}` followed by the text of the abstract. The abstract should normally be restricted to a single paragraph and is terminated by the command `\end{abstract}`

8.1. Sample coding for the start of an article

The code for the start of a title page of a typical paper might read:

```
\title{The anomalous magnetic moment of the  
neutrino and its relation to the solar neutrino problem}
```

```
\author{P J Smith1, T M Collins2,  
R J Jones3}\footnote[4]{Present address:  
Department of Physics, University of Bristol, Tyndalls Park Road,  
Bristol BS8 1TS, UK.} and Janet Williams3}
```

```
\address{$^1$ Mathematics Faculty, Open University,  
Milton Keynes MK7 6AA, UK}  
\address{$^2$ Department of Mathematics,  
Imperial College, Prince Consort Road, London SW7 2BZ, UK}  
\address{$^3$ Department of Computer Science,  
University College London, Gower Street, London WC1E 6BT, UK}
```

```
\ead{williams@ucl.ac.uk}
```

```
\begin{abstract}  
The abstract appears here.  
\end{abstract}
```

9. The text

The text of the article should be produced using standard L^AT_EX formatting. Articles may be divided into sections and subsections, but the length limit provided by the conference organizer should be adhered to.

9.1. Acknowledgments

Authors wishing to acknowledge assistance or encouragement from colleagues, special work by technical staff or financial support from organizations should do so in an unnumbered Acknowledgments section immediately following the last numbered section of the paper. The command `\ack` sets the acknowledgments heading as an unnumbered section.

9.2. Appendices

Technical detail that it is necessary to include, but that interrupts the flow of the article, may be consigned to an appendix. Any appendices should be included at the end of the main text of the paper, after the acknowledgments section (if any) but before the reference list. If there are

two or more appendices they will be called Appendix A, Appendix B, etc. Numbered equations will be in the form (A.1), (A.2), etc, figures will appear as figure A1, figure B1, etc and tables as table A1, table B1, etc.

The command `\appendix` is used to signify the start of the appendixes. Thereafter `\section`, `\subsection`, etc, will give headings appropriate for an appendix. To obtain a simple heading of ‘Appendix’ use the code `\section*{Appendix}`. If it contains numbered equations, figures or tables the command `\appendix` should precede it and `\setcounter{section}{1}` must follow it.

10. References

In the online version of *Journal of Physics: Conference Series* references will be linked to their original source or to the article within a secondary service such as INSPEC or ChemPort wherever possible. To facilitate this linking extra care should be taken when preparing reference lists.

Two different styles of referencing are in common use: the Harvard alphabetical system and the Vancouver numerical system. For *Journal of Physics: Conference Series*, the Vancouver numerical system is preferred but authors should use the Harvard alphabetical system if they wish to do so. In the numerical system references are numbered sequentially throughout the text within square brackets, like this [2], and one number can be used to designate several references.

10.1. Using BibT_EX

We highly recommend the `iopart-num` BibT_EX package by Mark A Caprio [1], which is included with this documentation.

10.2. Reference lists

A complete reference should provide the reader with enough information to locate the article concerned, whether published in print or electronic form, and should, depending on the type of reference, consist of:

- name(s) and initials;
- date published;
- title of journal, book or other publication;
- titles of journal articles may also be included (optional);
- volume number;
- editors, if any;
- town of publication and publisher in parentheses for *books*;
- the page numbers.

Up to ten authors may be given in a particular reference; where there are more than ten only the first should be given followed by ‘*et al*’. If an author is unsure of a particular journal’s abbreviated title it is best to leave the title in full. The terms *loc. cit.* and *ibid.* should not be used. Unpublished conferences and reports should generally not be included in the reference list and articles in the course of publication should be entered only if the journal of publication is known. A thesis submitted for a higher degree may be included in the reference list if it has not been superseded by a published paper and is available through a library; sufficient information should be given for it to be traced readily.

10.3. Formatting reference lists

Numeric reference lists should contain the references within an unnumbered section (such as `\section*{References}`). The reference list itself is started by the code `\begin{thebibliography}{<num>}`, where `<num>` is the largest number in the reference list and is completed by `\end{thebibliography}`. Each reference starts with `\bibitem{<label>}`, where ‘label’ is the label used for cross-referencing. Each `\bibitem` should only contain a reference to a single article (or a single article and a preprint reference to the same article). When one number actually covers a group of two or more references to different articles, `\nonum` should replace `\bibitem{<label>}` at the start of each reference in the group after the first.

For an alphabetic reference list use `\begin{thereferences} ... \end{thereferences}` instead of the ‘thebibliography’ environment and each reference can be start with just `\item` instead of `\bibitem{label}` as cross referencing is less useful for alphabetic references.

10.4. References to printed journal articles

A normal reference to a journal article contains three changes of font (see table 2) and is constructed as follows:

- the authors should be in the form surname (with only the first letter capitalized) followed by the initials with no periods after the initials. Authors should be separated by a comma except for the last two which should be separated by ‘and’ with no comma preceding it;
- the article title (if given) should be in lower case letters, except for an initial capital, and should follow the date;
- the journal title is in italic and is abbreviated. If a journal has several parts denoted by different letters the part letter should be inserted after the journal in Roman type, e.g. *Phys. Rev. A*;
- the volume number should be in bold type;
- both the initial and final page numbers should be given where possible. The final page number should be in the shortest possible form and separated from the initial page number by an en rule ‘–’, e.g. 1203–14, i.e. the numbers ‘12’ are not repeated.

A typical (numerical) reference list might begin

- [1] Strite S and Morkoc H 1992 *J. Vac. Sci. Technol.* B **10** 1237
- [2] Jain S C, Willander M, Narayan J and van Overstraeten R 2000 *J. Appl. Phys.* **87** 965
- [3] Nakamura S, Senoh M, Nagahama S, Iwase N, Yamada T, Matsushita T, Kiyoku H and Sugimoto Y 1996 *Japan. J. Appl. Phys.* **35** L74
- [4] Akasaki I, Sota S, Sakai H, Tanaka T, Koike M and Amano H 1996 *Electron. Lett.* **32** 1105
- [5] O’Leary S K, Foutz B E, Shur M S, Bhapkar U V and Eastman L F 1998 *J. Appl. Phys.* **83** 826
- [6] Jenkins D W and Dow J D 1989 *Phys. Rev. B* **39** 3317

which would be obtained by typing

```
\begin{thebibliography}{9}
\item Strite S and Morkoc H 1992 {\it J. Vac. Sci. Technol.} B {\bf 10} 1237
\item Jain S C, Willander M, Narayan J and van Overstraeten R 2000
{\it J. Appl. Phys.} {\bf 87} 965
\item Nakamura S, Senoh M, Nagahama S, Iwase N, Yamada T, Matsushita T, Kiyoku H
and Sugimoto Y 1996 {\it Japan. J. Appl. Phys.} {\bf 35} L74
\item Akasaki I, Sota S, Sakai H, Tanaka T, Koike M and Amano H 1996
{\it Electron. Lett.} {\bf 32} 1105
\item O’Leary S K, Foutz B E, Shur M S, Bhapkar U V and Eastman L F 1998
{\it J. Appl. Phys.} {\bf 83} 826
\item Jenkins D W and Dow J D 1989 {\it Phys. Rev.} B {\bf 39} 3317
\end{thebibliography}
```

Table 2. Font styles for a reference to a journal article.

Element	Style
Authors	Roman type
Date	Roman type
Article title (optional)	Roman type
Journal title	Italic type
Volume number	Bold type
Page numbers	Roman type

10.5. References to *Journal of Physics: Conference Series* articles

Each conference proceeding published in *Journal of Physics: Conference Series* will be a separate *volume*; references should follow the style for conventional printed journals. For example:

[1] Douglas G 2004 *J. Phys.: Conf. Series* **1** 23–36

10.6. References to preprints

For preprints there are two distinct cases:

- (1) Where the article has been published in a journal and the preprint is supplementary reference information. In this case it should be presented as:

[1] Kunze K 2003 T-duality and Penrose limits of spatially homogeneous and inhomogeneous cosmologies *Phys. Rev. D* **68** 063517 (*Preprint* gr-qc/0303038)

- (2) Where the only reference available is the preprint. In this case it should be presented as

[1] Milson R, Coley A, Pravda V and Pravdova A 2004 Alignment and algebraically special tensors *Preprint* gr-qc/0401010

10.7. References to electronic-only journals

In general article numbers are given, and no page ranges, as most electronic-only journals start each article on page 1.

- For *New Journal of Physics* (article number may have from one to three digits)
[1] Fischer R 2004 Bayesian group analysis of plasma-enhanced chemical vapour deposition data *New. J. Phys.* **6** 25
- For SISSA journals the volume is divided into monthly issues and these form part of the article number
[1] Horowitz G T and Maldacena J 2004 The black hole final state *J. High Energy Phys.* JHEP02(2004)008
[2] Bentivegna E, Bonanno A and Reuter M 2004 Confronting the IR fixed point cosmology with high-redshift observations *J. Cosmol. Astropart. Phys.* JCAP01(2004)001

10.8. References to books, conference proceedings and reports

References to books, proceedings and reports are similar to journal references, but have only two changes of font (see table 3).

Points to note are:

Table 3. Font styles for references to books, conference proceedings and reports.

Element	Style
Authors	Roman type
Date	Roman type
Book title (optional)	Italic type
Editors	Roman type
Place (city, town etc) of publication	Roman type
Publisher	Roman type
Volume	Roman type
Page numbers	Roman type

- Book titles are in italic and should be spelt out in full with initial capital letters for all except minor words. Words such as Proceedings, Symposium, International, Conference, Second, etc should be abbreviated to *Proc.*, *Symp.*, *Int.*, *Conf.*, *2nd*, respectively, but the rest of the title should be given in full, followed by the date of the conference and the town or city where the conference was held. For Laboratory Reports the Laboratory should be spelt out wherever possible, e.g. *Argonne National Laboratory Report*.
- The volume number, for example vol 2, should be followed by the editors, if any, in a form such as ‘ed A J Smith and P R Jones’. Use *et al* if there are more than two editors. Next comes the town of publication and publisher, within brackets and separated by a colon, and finally the page numbers preceded by p if only one number is given or pp if both the initial and final numbers are given.

Examples taken from published papers:

- [1] Kurata M 1982 *Numerical Analysis for Semiconductor Devices* (Lexington, MA: Heath)
- [2] Selberherr S 1984 *Analysis and Simulation of Semiconductor Devices* (Berlin: Springer)
- [3] Sze S M 1969 *Physics of Semiconductor Devices* (New York: Wiley-Interscience)
- [4] Dorman L I 1975 *Variations of Galactic Cosmic Rays* (Moscow: Moscow State University Press) p 103
- [5] Caplar R and Kulisic P 1973 *Proc. Int. Conf. on Nuclear Physics (Munich)* vol 1 (Amsterdam: North-Holland/American Elsevier) p 517
- [6] Cheng G X 2001 *Raman and Brillouin Scattering-Principles and Applications* (Beijing: Scientific)
- [7] Szytula A and Leciejewicz J 1989 *Handbook on the Physics and Chemistry of Rare Earths* vol 12, ed K A Gschneidner Jr and L Erwin (Amsterdam: Elsevier) p 133
- [8] Kuhn T 1998 *Density matrix theory of coherent ultrafast dynamics Theory of Transport Properties of Semiconductor Nanostructures* (Electronic Materials vol 4) ed E Schöll (London: Chapman and Hall) chapter 6 pp 173–214

11. Tables and table captions

Tables should be numbered serially and referred to in the text by number (table 1, etc, **rather than** tab. 1). Each table should be a float and be positioned within the text at the most convenient place near to where it is first mentioned in the text. It should have an explanatory caption which should be as concise as possible.

11.1. The basic table format

The standard form for a table is:

```
\begin{table}
\caption{\label{label}Table caption.}
\begin{center}
```

```

\begin{tabular}{llll}
\br
Head 1&Head 2&Head 3&Head 4\\
\mr
1.1&1.2&1.3&1.4\\
2.1&2.2&2.3&2.4\\
\br
\end{tabular}
\end{center}
\end{table}

```

The above code produces table 4.

Table 4. Table caption.

Head 1	Head 2	Head 3	Head 4
1.1	1.2	1.3	1.4
2.1	2.2	2.3	2.4

Points to note are:

- (1) The caption comes before the table.
- (2) The normal style is for tables to be centred in the same way as equations. This is accomplished by using `\begin{center} ... \end{center}`.
- (3) The default alignment of columns should be aligned left.
- (4) Tables should have only horizontal rules and no vertical ones. The rules at the top and bottom are thicker than internal rules and are set with `\br` (bold rule). The rule separating the headings from the entries is set with `\mr` (medium rule). These commands do not need a following double backslash.
- (5) Numbers in columns should be aligned as appropriate, usually on the decimal point; to help do this a control sequence `\lineup` has been defined which sets `\0` equal to a space the size of a digit, `\m` to be a space the width of a minus sign, and `\-` to be a left overlapping minus sign. `\-` is for use in text mode while the other two commands may be used in maths or text. (`\lineup` should only be used within a table environment after the caption so that `\-` has its normal meaning elsewhere.) See table 5 for an example of a table where `\lineup` has been used.

12. Figures and figure captions

Figures must be included in the source code of an article at the appropriate place in the text not grouped together at the end.

Each figure should have a brief caption describing it and, if necessary, interpreting the various lines and symbols on the figure. As much lettering as possible should be removed from the figure itself and included in the caption. If a figure has parts, these should be labelled *(a)*, *(b)*, *(c)*, etc. Table 6 gives the definitions for describing symbols and lines often used within figure captions (more symbols are available when using the optional packages loading the AMS extension fonts).

Authors should try and use the space allocated to them as economically as possible. At times it may be convenient to put two figures side by side or the caption at the side of a figure.

Table 5. A simple example produced using the standard table commands and `\lineup` to assist in aligning columns on the decimal point. The width of the table and rules is set automatically by the preamble.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>	<i>F</i>	<i>G</i>
23.5	60	0.53	−20.2	−0.22	1.7	14.5
39.7	−60	0.74	−51.9	−0.208	47.2	146
123.7	0	0.75	−57.2	—	—	—
3241.56	60	0.60	−48.1	−0.29	41	15

Table 6. Control sequences to describe lines and symbols in figure captions.

Control sequence	Output	Control sequence	Output
<code>\dotted</code>	<code>\opencircle</code>	○
<code>\dashed</code>	----	<code>\opentriangle</code>	△
<code>\broken</code>	---	<code>\opentriangledown</code>	▽
<code>\longbroken</code>	— — —	<code>\fullsquare</code>	■
<code>\chain</code>	— . —	<code>\opensquare</code>	□
<code>\dashddot</code>	— . . —	<code>\fullcircle</code>	●
<code>\full</code>	——	<code>\opendiamond</code>	◇

To put figures side by side, within a figure environment, put each figure and its caption into a minipage with an appropriate width (e.g. 3in or 18pc if the figures are of equal size) and then separate the figures slightly by adding some horizontal space between the two minipages (e.g. `\hspace{.2in}` or `\hspace{1.5pc}`). To get the caption at the side of the figure add the small horizontal space after the `\includegraphics` command and then put the `\caption` within a minipage of the appropriate width aligned bottom, i.e. `\begin{minipage}[b]{3in}` etc (see code in this file used to generate figures 1–3).

Note that it may be necessary to adjust the size of the figures (using optional arguments to `\includegraphics`, for instance `[width=3in]`) to get you article to fit within your page allowance or to obtain good page breaks.

Using the `graphicx` package figures can be included using code such as:

```
\begin{figure}
\begin{center}
\includegraphics{file.eps}
\end{center}
\caption{\label{label}Figure caption}
\end{figure}
```

References

- [1] IOP Publishing is to grateful Mark A Caprio, Center for Theoretical Physics, Yale University, for permission to include the `iopart-num` Bib_{La}T_EXpackage (version 2.0, December 21, 2006) with this documentation. Updates and new releases of `iopart-num` can be found on www.ctan.org (CTAN).

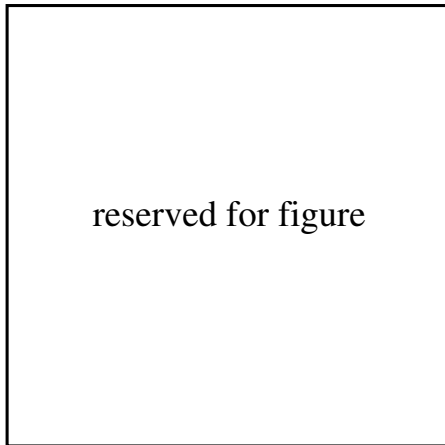


Figure 1. Figure caption for first of two sided figures.

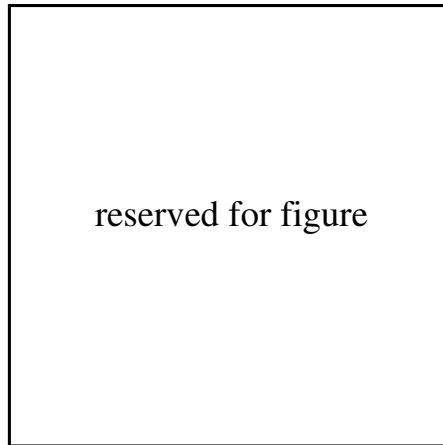


Figure 2. Figure caption for second of two sided figures.

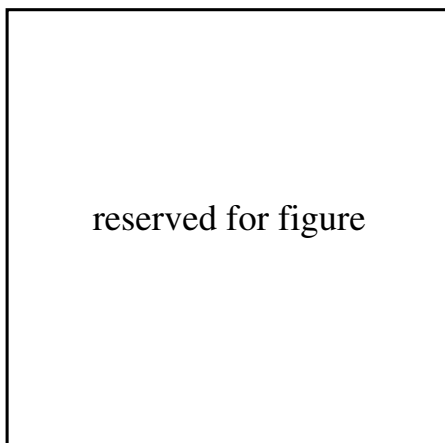


Figure 3. Figure caption for a narrow figure where the caption is put at the side of the figure.