Using a Two-Layer Neural Network and Physicochemical Properties to Classify Glass for Forensic Analysis

Michael Arango, Mark Barna & Paul Brewster

June 24, 2017

1 Project Proposal

In this study, we will use the physicochemical properties of glass to determine whether or not a given glass sample was taken from a window. This is a fundamental problem in forensic analysis as it is highly unlikely that glass fragments will be found on people unless they have been present at the time glass breaks. Glass analysis is of vital importance in forensic science as it allows us to test if the glass fragment found on a person is the same as the glass at a crime scene. Since glass is made up of several raw materials and certain elements impart specific properties, we can find out a lot about the glass if we analyze the chemical composition.

The dataset we chose for analysis was made available for download from the UCI Machine Learning Repository and was created by the USA Forensic Science Service. There are 214 observations of 9 different features along with 214 targets that specify whether the glass sample came from a window or not. While we would like more data to train a neural network, we believe the dataset is large enough for our purposes. It is difficult to know before we train a neural network if we have enough data, but the amount of data required is directly related to the complexity of the underlying decision boundary we are trying to implement. We won't know how complex the decision boundary we are trying to approximate is until we train the network, but we feel confident using the dataset as many others have used the dataset and found robust results. Several others papers in the literature use much more complex methods than we will employ and have not found the size of the data to be an issue.

We have chosen a two-layer perceptron network with tangent-sigmoid transfer functions in the hidden layer and *softmax* transfer functions in the output layer. This is a

fairly standard network for pattern recognition. Moreover, we will use the *Scaled Conjugate Gradient (SGD)* algorithm to train the network as it is good for pattern recognition problems in which the output layer uses a non-linear transfer function. Since we do not expect the training error to converge to zero, we implement early stopping criteria to prevent overfitting. Lastly, we use *cross-entropy* as our performance index since our targets take on discrete values and it is the optimal performance index for pattern recognition networks that use the *softmax* transfer function in the output layer.

Two different frameworks will be used to implement the neural network. First, we will use the Neural Network Toolbox, specifically the Neural Network Pattern Recognition Tool (`nprtool`) train, validate, and test our network. We use this framework to start with a simple graphical user interface to quickly ensure our specified network architecture is appropriate and to get baseline performance statistics. Then, we will replicate the analysis in Python to gain practical experience building network architectures in a scripting language. Note that since the goal is practical experience, we will not be leveraging the power of the *scikit-learn* package (`sklearn`) for this exercise.

Several reference materials will be consulted to obtain sufficient background knowledge of the subject at hand. First, we plan on doing a thorough review of the forensic chemistry and geology literature to understand the reasons for using physicochemical properties to classify glass. Then, papers on glass analysis will be examined to supplement background knowledge with experiential knowledge.

Considering our problem is one of pattern recognition, a confusion matrix will be used to assess the accuracy of our model and the *false postive* (Type I error) and *false negative* (Type II error) rates. Further, the *Receiver Operating Characteristic (ROC) curve* will be used to compare the true positive rate to the false positive rate. This will help us gain additional knowledge of the predictive power of our network.

We plan to finish our research and submit it by Wednesday, June 28, 2017.

2 Literature Review

3 Preparing your manuscript

3.1 General guidelines

1. LaTeX and $\mathcal{A}\mathcal{M}\mathcal{S}$-LaTeX provide a rich set of commands for all common, important features of your paper. Use them; avoid definitions and use of custom commands.

2. There is no need to redefine any TeX, LaTeX or $\mathcal{A}\mathcal{M}\mathcal{S}$-LaTeX commands.

3. Avoid direct formatting for headings cleanly set as section headings.

4. Use LaTeX commands for font changes. For example: use `\textbf{phrase}`, not `{\bf phrase}`; use `\mathcal{C}`, not `{\cal C}`; etc.

3.2 How to start with `ouparticle.cls`

Before you type anything that actually appears in the paper, you need to include a `\documentclass{ouparticle}` command at the very beginning, and then the two commands that have to be part of any LaTeX document, `\begin{document}` at the start and `\end{document}` at the end of your paper.

3.3 Document structure

The main structure of your paper is as follows:

```
\documentclass[12pt,...]{ouparticle}
\usepackage[...]{packages}

\title{...}
\author{
    \name{...}
    \address{...}
    \email{...}
        \and
    \name{...}
    \address{...}
    \email{...}
        \and
    \name{...}
    \address{...}
    \email{...}
}
\abstract{...}
\keywords{...}

\maketitle

\begin{document}

\section{....}
...
```

```
\subsection{....}
....
\end{document}
```

3.4 Options

By default, all of the options within `article.cls` are available with this class file. This class file provides the following additional options.

oneline: This option will set your entire manuscript in one line spacing. It will not affect the footnote, figure and table environments.

halfline: This is to set your entire manuscript in half line spacing.

endnotes: To make all footnotes to endnotes. You may follow the same coding `\footnote{text}` for both footnotes and endnotes. Once you use this option you have to use the `\theendnotes` command at the place where all the endnotes have to be set in your paper.

numbib: This is the default option that numbers the bibliography items; this option does nothing with natbib and other packages.

nonumbib: For unnumbered bibliography.

3.5 Front matter

The title of the manuscript is simply specified by using the `\title{text}` command in the same manner as in this sample. Author's information consists of the name of the author and the corresponding institutions with addresses, as given in this example. Include an electronic mail address if available, inserting it into the `\email{text}` commands. You may follow the same coding if there are more than one author; separate authors with `\and`. Please identify the corresponding author with his/her electronic mail address by `\thanks{text}`. An abstract for your paper is specified by using `\abstract{text}`. A `\keywords{text}` macro may also be used to indicate keywords for the article. Use `\maketitle` after the abstract and keywords to make the header of your article.

3.6 Sections and subsections

To begin a new section, give the heading of that section in the `\section{text}` command. A section number is supplied automatically. Use the starred form (`\section*{text}`) of the command to suppress the automatic numbering. If you want to be able to make reference to that section, then you need to `label` it (see Section 3.14). You can

have sections up to five levels. The sectioning commands are `\section`, `\subsection`, `\subsubsection`, `\paragraph` and `\subparagraph`.

3.7 Ordinary text

The ends of words and sentences are marked by spaces. It does not matter how many spaces you type. The end of a line counts as a space. One or more blank lines denote the end of a paragraph.

There are a number of things for which you need to follow different methods. As you know, quotation marks, quotes within quotes, dashes, ellipsis, etc. should be as per the LaTeX standard input. LaTeX interprets some common characters as commands, and therefore you must instead type those common characters as specific LaTeX commands to generate them. Those characters are $, &, %, #, {, and }.

3.8 Formatting

One should always use LaTeX macros rather than the lower-level TeX macros like `\it`, `\bf` and `\tt`. The LaTeX macros offer much improved features. The following table summarizes the font selection commands in LaTeX.

LaTeX text formatting commands

`\textit`	Italics	`\textsf`	Sans Serif
`\textbf`	Boldface	`\textsc`	Small Caps
`\texttt`	Typewriter	`\textmd`	Medium Series
`\textrm`	Roman	`\textnormal`	Normal Series
`\textsl`	Slanted	`\textup`	Upright Series

LaTeX math formatting commands

`\mathit`	Math Italics	`\mathfrak`	Fraktur
`\mathbf`	Math Boldface	`\mathbb`	Blackboard Bold
`\mathtt`	Math Typewriter	`\mathnormal`	Math Normal
`\mathsf`	Math Sans Serif	`\boldsymbol`	Bold math for Greek letters
`\mathcal`	Calligraphic		and other symbols

3.9 Figures and tables

Use normal LaTeX coding for figures and tables. Figure and table environments should be inserted after (not in) the paragraph in which the figure is first mentioned or grouped all together at the end of the file. They will be numbered automatically. The following is an example of typesetting a table.

```
\begin{table}
\caption{Table caption text.}
\label{key}
The table matter goes here.
\end{table}
```

As always with LaTeX, the `\label` must be after the `\caption`, and inside the figure or table environment. The reference for figures and tables inside text can be made using the `\ref{key}` command.

3.10 Equations

Equations are used in the same way as described in the LaTeX manual. Do not start a paragraph with a displayed equation. Equations are numbered consecutively, with equation numbers in parentheses flush right. For example, if you type

```
\begin{equation}\label{eq1}
\int^{r_2}_0 F(r,\varphi){\rm d}r\,{\rm d}\varphi = [\sigma r_2/(2\mu_0)]
\int^{\infty}_0\exp(-\lambda|z_j-z_i|)\lambda^{-1}J_1 (\lambda r_2)J_0
(\lambda r_i\,\lambda {\rm d}\lambda)
\end{equation}
```

then you will get the following output:

$$\int_0^{r_2} F(r,\varphi)\mathrm{d}r\,\mathrm{d}\varphi = [\sigma r_2/(2\mu_0)] \int_0^{\infty} \exp(-\lambda|z_j - z_i|)\lambda^{-1} J_1(\lambda r_2) J_0(\lambda r_i\,\lambda \mathrm{d}\lambda) \qquad (1)$$

It inserts space both above and below the equation. $\mathcal{AMS}$-LaTeX has several environments that make it easier to typeset complicated multiline displayed equations. These are explained in the $\mathcal{AMS}$-LaTeX User Guide. A `subequation` environment is available to create equations with sub-numbering of the equation counter. It takes one (optional) argument to specify the way that the sub-counter should appear.

3.11 Displayed text

Text is displayed by indenting it from the left and right margins. Quotations are commonly displayed. There are short quotations:

> This is a short quotation. It consists of a single paragraph of text. See how it is formatted.

and longer ones:

> This is a longer quotation. It consists of two paragraphs of text, neither of which are particularly interesting.
>
> This is the second paragraph of the quotation. It is just as dull as the first paragraph.

You can even display poetry.

> There is an environment for verse
> Whose features some poets will curse.
>
> For instead of making
> Them do *all* line breaking,
> It allows them to put too many words on a line when they'd rather be forced to be terse.

3.12 Listings

Another frequently displayed structure is a list. The following is an example of an *itemized* list.

- This is the first item of an itemized list. Each item in the list is marked with a '•'.
- This is the second item of the list. It contains another list nested inside it. The inner list is an *enumerated* list.
  1. This is the first item of an enumerated list that is nested within the itemized list.
  2. This is the second item of the inner list. LaTeX allows you to nest lists deeper than you really should.

  This is the rest of the second item of the outer list. It is no more interesting than any other part of the item.
- This is the third item of the list.

3.13 Displayed sentences: theorems and such

These environments have to be defined with the help of LaTeX's `\newtheorem` command, and also with the $\mathcal{AMS}$-LaTeX package for theorems that is already with your class file. For example, `\newtheorem{thm}{Theorem}`. Predefined theorem styles can be used in your article to differentiate the theorem-like environments. You can have an extra command, `\newproof`, that can be used for displayed text. The following is an example of using the above-defined `thm` environment.

```
\begin{thm}
This is body matter for this environment.
\end{thm}
```

3.14 Cross-referencing

LaTeX possesses features for labelling and cross-referencing section headings, equations, tables, figures and theorems. Their proper usage in the context of section headings, equations, tables and figures are discussed in the appropriate sections.

Cross-referencing depends upon the use of 'keys' that are defined by the user. The `\label{key}` command is used to identify the links. Keys are strings of characters that serve to label section headings, equations, tables and figures that replace explicit, by-hand numbering. The `\ref{key}` command is used for cross-referencing.

Files that use cross-referencing (and almost all manuscripts do) need to be processed through LaTeX at least twice to ensure that the keys have been properly linked to the appropriate numbers.

3.15 Footnotes and endnotes

The footnote text can either appear at the bottom of a page or at the end of your paper. The `\footnote` macro *should not* be used in the front matter to provide additional information about authors (such as corresponding addresses); instead, use `\thanks{text}` commands. The document option '`endnotes`' is used to make endnotes. The command `\theendnotes` should be used to place the endnotes at the required location in the text. They will be put in a separate 'Notes' section.

3.16 Appendix

The `\appendix` command signals that all following sections are appendices, and therefore the headings after `\appendix` will be set as appendix headings. For a single appendix, use `\appendix*` followed by the `\section{text}` command to suppress the appendix letter in the section heading.

3.17 Special sections for notes and acknowledgements

If you wish to include a 'Notes' or 'Acknowledgements' section in your paper, use the `\begin{notes}...\end{notes}` macro. We use the same environment for both 'Notes' and 'Acknowledgements'. The following examples show to how to use this macro.

```
\begin{notes}
Please note that this class file is provided as it is, and
```

```
copyright by Oxford University Press. You are free to use this
class file, provided that you do not make changes in this class file.
If you do make changes, you are requested to rename the class file.
\end{notes}

\begin{notes}[Acknowledgements]
The authors would like to thank...
\end{notes}
```

3.18 References

The reference entries can be LaTeX typed bibliographies or generated through a BIBTeX database. BIBTeX is an adjunct to LaTeX that aids in the preparation of bibliographies. BIBTeX allows authors to build up a database or collection of bibliography entries that may be used for many manuscripts. They also save us the trouble of having to specify formatting. More details can be found in the *BIBTeX Guide*. For LaTeX reference entries use the `\begin{thebibliography}....\end{thebibliography}` environment (see below) to make references in your paper. We have provided the class file option to distinguish two styles of references. Those options are `numbib` and `nonumbib`. You can select one of these options with the `\documentclass` command. By default the class file will take the `numbib` option. The following is an example of LaTeX bibliography.

```
\begin{thebibliography}{0}
\bibitem{bib1}
Goossens, M., F. Mittelbach, and A. Samarin: {\em The {\LaTeX} Companion}.
Addison-Wesley, Reading, MA, USA, 1994.
\bibitem{bib2}
Knuth, D.E: {\em The {\TeX}book}. Addison-Wesley, Reading, MA, USA, 1984.
\bibitem{bib3}
Lamport, L.: {\em {\LaTeX} -- A Document Preparation System -- User's
Guide and Reference Manual}. Addison-Wesley, Reading, MA, USA, 1985.
\bibitem{bib4}
Smith, I.N., R.S. Johnes, and W.P. Hines: 1992, 'Title of the Article',
\textit{Journal Title in Italics} \textbf{Vol. no. X}, pp. 00--00
\end{thebibliography}
```

4 Macro packages

The following packages are compulsorily needed by the class file:

```
amsmath         graphicx
amssymb         endnotes
amsfonts        setspace
verbatim        geometry
```

The commonly used packages already used by this class file that authors can use whenever required are:

```
xspace          latexsym        url
amscd           multicol        algorithm
rotating        array           subfigure
```

Additionally, you can use other packages and these should be loaded using the `\usepackage` command.