CPSC 66 Final Report: Examples and Requirements

Student 1 Name USERID1@SWARTHMORE.EDU
Student 2 Name USERID1@SWARTHMORE.EDU

Abstract

This paper contains examples and instructions for your final report. This is also an example for how to use the templates to write your paper in LaTeX. **This document is not authoritative, you should use the course Project Writeup for the exact requirements of your final project**.

This template is a modification of the template used for official publications in the annual International Conference on Machine Learning (ICML 2014)[1]. Using this template will make conforming to page formatting requirements trivial – there is no extra work required to get the correct font size, margins, spacing, and font types. Submissions must be compiled as a PDF (you can use the provided Makefile to do this).

The most important thing to keep in mind as you write your paper: "who is my intended audience?" This is true for course projects as well as submission of research. For this report, your intended audience is your fellow classmate. That is, as you explain your work, you should assume the reader has a broad understanding of the field of machine learning, but does not have familiarity with your specific project. It is important to motivate and explain your approach in this context. We highly recommend having a classmate proofread your paper at multiple checkpoints.

1. Introduction

All papers should have, at a minimum (a more complete description is available on the Project Writeup):

- A one- or two-paragraph abstract that outlines the central goal and results of the project. This is your 30-second elevator pitch where you sell a reader on reading your paper. It should be 200 words maximum.
- A descriptive title
- Bibliographic references.
- Several figures – especially for your methods and results
- An **Introduction** – what you attempted to do and what was the motivation for your work. You should provide some context about the problem including any relevant background about the task and related work.
- **Approach** or **Methods** – what you did. Explain the algorithms you used to solve the problem. You should not talk about your code or implementation details - we can look at that for details. Rather, use figures and pseudocode, as in Algorithm 1 to explain your approach and the major concepts. See the example papers provided.
- **Experiments and Results** – how did you evaluate the quality of your approach? For most of you this is empirical (experiments). But you can also include analytical (e.g., run-time analysis). You should explain your **experimental methodology** and your **results**. Methodology includes: explain the data set(s) you use for evaluation. Where did you obtain them? Did you do any significant pre-processing (throw away noisy/corrupted data; generate features; fill in missing values; aggregate fields, normalize features)? Did you employ cross-validation? This should explain how I could re-produce your experiments. Results should include charts and figures. This section requires the most effort as your goal is to supply both a *rich* and *concise* explanation. Merely listing a bunch of raw numbers will not impress any audience.
- **Discussion** – what are the implications of your approach? This can be included with the evaluation. E.g., have a **Experimental Methodology** and **Experimental Results** section. What are the lessons from

[1] I find it aesthetically pleasing and easy to u...woah! I created a footnote!

CPSC 66 Machine Learning Proceedings, Swarthmore College, Spring 2021.

your experiments? If they failed, why? What are the limitations of your approach? The advantages? What future work can you suggest? How does this compare to related work (cited papers or course discussion). Your paper should have a **Conclusion** section where these last few questions can be answered.

A note about expectations: failure is a part of the scientific process. We are not expecting you to produce state-of-the-art discoveries. If your experiments do not turn out as you expect, you are in the same situation every scientist has been in many times in their career. The importance of this project is to measure how well you carried out the process. You will be graded on (i) how clearly you defined the problem (your hypothesis, objective, and motivation) and your approach, (ii) the appropriateness of your experiments in addressing your hypothesis/objectives, and (iii) the quality of your report on your results.

2. Getting started

Your main directory has a directory for `code` (see the online requirements for this directory) and a directory `paper`. Your final project paper should go in this later directory.

2.1. Provided Files

You have been provided with the following files in the directory `paper/example`:

- `Makefile` – to aid in compiling your document. You should not need to modify this except **you must change the target** at the top of the file. The target should to be the name of your paper (your userids e.g., asoni1-soni).
- `example.tex` – the source file for this document. You should use this file as a guide for your main source file.
- `example.bib` – an example bibliography file. Your referenced work should be formatted in a similar file with the same name as your source tex file.
- `icml2014.{sty,bst}` – style files for formatting your paper. Do not modify these; it will cause your paper to not compile correctly when you submit it.
- `icml_numpapers.pdf` – example image used in this paper. Your images do not need to be pdfs – png, jpeg, or eps should work without problems.

Do not use the example directory as a working directory. Instead, copy any files you want up to `paper` or start from the empty files I have given.

2.2. Compiling Document

The easiest way to turn this LaTeX into a PDF, is to use the `Makefile` in your project directory. The `Makefile` will compile your file (and your bibliography file) and turn it into a PDF.

Here are the basic instructions:

- `make` will create a PDF file from your LaTeX document.
- `make view` will display the PDF file (on lab machines only; type `open paper.pdf` on a Mac).
- `make clean` will clean up some files you might not need
- `make cleanall` will clean up all non-source files

Change the target to the name of your final submission (see Section 3.1).

3. Paper Requirements

Note that using this example file should automatically meet the formatting requirements. But to be complete, review this section.

The formatting instructions below will be enforced when evaluating your report. Please contact me early if there are any issues.

- The maximum paper length is 6 pages plus references.
- Do not alter the style template; in particular, do not compress the paper format by reducing the vertical spaces.
- Place figure captions *under* the figure (and omit titles from inside the graphic file itself). Place table captions *over* the table. All axes should be labeled on graphs. Captions should be self-sufficient – I should be able to interpret the image without having to refer to the text for simple explanations.
- References must include page numbers whenever possible and be as complete as possible.

3.1. Submitting Papers

Submissions will be entirely electronic, via your `Project-userID1-userID2` Github repository. See the syllabus Project Writeup for details.

3.2. Length and Dimensions

Papers should be at most 6 pages, including all figures, tables, and appendices.

NOTE: what follows is the boilerplate from the ICML conference and is included more to explain how to use the style file; nothing should be interpreted as a requirement for your project.

The text of the paper should be formatted in two columns, with an overall width of 6.75 inches, height of 9.0 inches, and 0.25 inches between the columns. The left margin should be 0.75 inches and the top margin 1.0 inch (2.54 cm). The right and bottom margins will depend on whether you print on US letter or A4 paper, but all final versions must be produced for US letter size.

The paper body should be set in 10 point type with a vertical spacing of 11 points. Please use Times typeface throughout the text.

3.3. Title

The paper title should be set in 14 point bold type and centered between two horizontal rules that are 1 point thick, with 1.0 inch between the top rule and the top edge of the page. Capitalize the first letter of content words and put the rest of the title in lower case.

3.4. Abstract

The paper abstract should begin in the left column, 0.4 inches below the final address. The heading 'Abstract' should be centered, bold, and in 11 point type. The abstract body should use 10 point type, with a vertical spacing of 11 points, and should be indented 0.25 inches more than normal on left-hand and right-hand margins. Insert 0.4 inches of blank space after the body. Keep your abstract brief and self-contained, limiting it to one paragraph and no more than six or seven sentences.

3.5. Partitioning the Text

You should organize your paper into sections and paragraphs to help readers place a structure on the material and understand its contributions.

3.5.1. Sections and Subsections

Section headings should be numbered, flush left, and set in 11 pt bold type with the content words capitalized. Leave 0.25 inches of space before the heading and 0.15 inches after the heading.

Similarly, subsection headings should be numbered, flush left, and set in 10 pt bold type with the content words capitalized. Leave 0.2 inches of space before the heading and 0.13 inches afterward.

Finally, subsubsection headings should be numbered, flush left, and set in 10 pt small caps with the content words capitalized. Leave 0.18 inches of space before the heading and 0.1 inches after the heading.

Please use no more than three levels of headings.

3.5.2. Paragraphs and Footnotes

Within each section or subsection, you should further partition the paper into paragraphs. Do not indent the first line of a given paragraph, but insert a blank line between succeeding ones.

You can use footnotes[2] to provide readers with additional information about a topic without interrupting the flow of the paper. Indicate footnotes with a number in the text where the point is most relevant. Place the footnote in 9 point type at the bottom of the column in which it appears. Precede the first footnote in a column with a horizontal rule of 0.8 inches.[3]

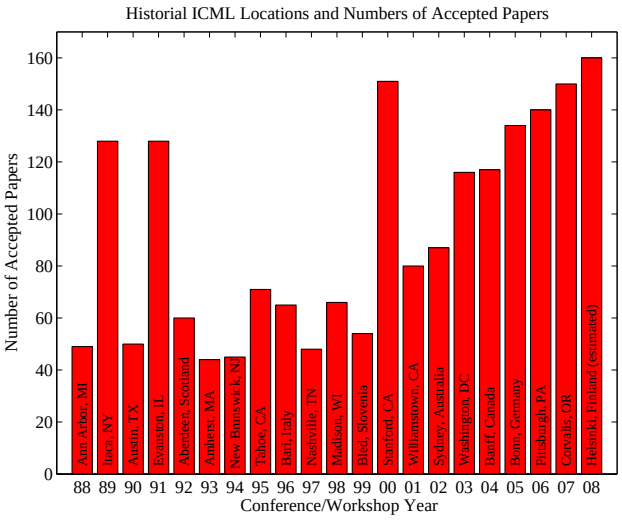

Figure 1. Historical locations and number of accepted papers for International Machine Learning Conferences (ICML 1993 – ICML 2008) and International Workshops on Machine Learning (ML 1988 – ML 1992). At the time this figure was produced, the number of accepted papers for ICML 2008 was unknown and instead estimated.

[2]For the sake of readability, footnotes should be complete sentences.

[3]Multiple footnotes can appear in each column, in the same order as they appear in the text, but spread them across columns and pages if possible.

Algorithm 1 Bubble Sort

Input: data x_i, size m
repeat
  Initialize $noChange = true$.
  for $i = 1$ **to** $m - 1$ **do**
    if $x_i > x_{i+1}$ **then**
      Swap x_i and x_{i+1}
      $noChange = false$
    end if
  end for
until $noChange$ is $true$

3.6. Figures

You may want to include figures in the paper to help readers visualize your approach and your results. Such artwork should be centered, legible, and separated from the text. Lines should be dark and at least 0.5 points thick for purposes of reproduction, and text should not appear on a gray background. Figures should be placed near their place of reference, rather than at the end of the document.

Label all distinct components of each figure. If the figure takes the form of a graph, then give a name for each axis and include a legend that briefly describes each curve. Do not include a title inside the figure; instead, the caption should serve this function.

Number figures sequentially, placing the figure number and caption *after* the graphics, with at least 0.1 inches of space before the caption and 0.1 inches after it, as in Figure 1. The figure caption should be set in 9 point type and centered unless it runs two or more lines, in which case it should be flush left. You may float figures to the top or bottom of a column, and you may set wide figures across both columns (use the environment `figure*` in LaTeX), but always place two-column figures at the top or bottom of the page.

3.7. Algorithms

If you are using LaTeX, please use the "algorithm" and "algorithmic" environments to format pseudocode. These require the corresponding stylefiles, algorithm.sty and algorithmic.sty, which are supplied with this package. Algorithm 1 shows an example.

3.8. Tables

You may also want to include tables that summarize material. Like figures, these should be centered, legible, and numbered consecutively. However, place the title *above* the table with at least 0.1 inches of space before the title and the same after it, as in Table 1. The table title should be set in 9 point type and centered unless it runs two or more lines, in which case it should be flush left.

Table 1. Classification accuracies for naive Bayes and flexible Bayes on various data sets.

DATA SET	NAIVE	FLEXIBLE	BETTER?
BREAST	95.9± 0.2	96.7± 0.2	√
CLEVELAND	83.3± 0.6	80.0± 0.6	×
GLASS2	61.9± 1.4	83.8± 0.7	√
CREDIT	74.8± 0.5	78.3± 0.6	
HORSE	73.3± 0.9	69.7± 1.0	×
META	67.1± 0.6	76.5± 0.5	√
PIMA	75.1± 0.6	73.9± 0.5	
VEHICLE	44.9± 0.6	61.5± 0.4	√

Tables contain textual material that can be typeset, as contrasted with figures, which contain graphical material that must be drawn. Specify the contents of each row and column in the table's topmost row. Again, you may float tables to a column's top or bottom, and set wide tables across both columns, but place two-column tables at the top or bottom of the page.

3.9. Equations

Your paper should contain several equations to define any mathematical terms. For example, you may want to define a voting scheme for k-nearest neighbors for a test example x_i with nearest neighbors knn:

$$p_i = \arg\max_c \sum_{\{x_j, y_j\} \in knn} I(y_j = c) \times \frac{1}{1 + dist(x_j, x_i)} \tag{1}$$

or the Silhouette Index S of a clustering algorithm:

$$S = \frac{1}{K} \sum_{j=1}^{K} \frac{1}{|c_j|} \sum_{x_i \in c_j} s(x_i) \tag{2}$$

3.10. Citations and References

Please use APA reference format regardless of your formatter or word processor. If you rely on the LaTeX bibliographic facility, use `natbib.sty` and `icml2014.bst` included in the style-file package to obtain this format.

Citations within the text should include the authors' last names and year. If the authors' names are included in the sentence, place only the year in parentheses, for example when referencing Arthur Samuel's pioneering work (1959). Otherwise place the entire reference in parentheses with the authors and year separated by a comma (Samuel, 1959).

List multiple references separated by semicolons (Kearns, 1989; Samuel, 1959; Mitchell, 1980). Use the 'et al.' construct only for citations with three or more authors or after listing all authors to a publication in an earlier reference (Michalski et al., 1983).

Authors should cite their own work in the third person in the initial version of their paper submitted for blind review. Please refer to Section **??** for detailed instructions on how to cite your own papers.

Use an unnumbered first-level section heading for the references, and use a hanging indent style, with the first line of the reference flush against the left margin and subsequent lines indented by 10 points. The references at the end of this document give examples for journal articles (Samuel, 1959), conference publications (Langley, 2000), book chapters (Newell & Rosenbloom, 1981), books (Duda et al., 2000), edited volumes (Michalski et al., 1983), technical reports (Mitchell, 1980), and dissertations (Kearns, 1989).

Alphabetize references by the surnames of the first authors, with single author entries preceding multiple author entries. Order references for the same authors by year of publication, with the earliest first. Make sure that each reference includes all relevant information (e.g., page numbers).

Acknowledgments

Place acknowledgements in an unnumbered section at the end of the paper. Typically, this will include to colleagues who contributed to the ideas, individuals who reviewed your submission, or external sources who helped acquire data.

References

Duda, R. O., Hart, P. E., and Stork, D. G. *Pattern Classification*. John Wiley and Sons, 2nd edition, 2000.

Kearns, M. J. *Computational Complexity of Machine Learning*. PhD thesis, Department of Computer Science, Harvard University, 1989.

Langley, P. Crafting papers on machine learning. In Langley, Pat (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (eds.). *Machine Learning: An Artificial Intelligence Approach, Vol. I*. Tioga, Palo Alto, CA, 1983.

Mitchell, T. M. The need for biases in learning generalizations. Technical report, Computer Science Department, Rutgers University, New Brunswick, MA, 1980.

Newell, A. and Rosenbloom, P. S. Mechanisms of skill acquisition and the law of practice. In Anderson, J. R. (ed.), *Cognitive Skills and Their Acquisition*, chapter 1, pp. 1–51. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1981.

Samuel, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):211–229, 1959.