
MLP Coursework 1

sXXXXXXX

Abstract

The abstract should be 100–200 words long, providing a concise summary of the contents of your report.

1. Introduction

This document provides a template for the MLP coursework 1 report. The template structures the report into sections, which you are recommended to use, but can change if you wish. If you want to use subsections within a section that is fine, but please do not use any deeper structuring. In this template the text in each section will include an outline of what you should include in each section, along with some practical LaTeX examples (for example figures, tables, algorithms). Your document must be no longer than **five pages**, with an additional page (or more!) allowed for references.

The introduction should place your work in context, giving the overall motivation for the work, and clearly outlining the research questions you have explored – in this case questions arising from an investigation of shallow and deep neural network models, hyper-parameter variations, and regularization/weight penalty. This section should also include a concise description of the Balanced EMNIST task and data – be precise: for example state the size of the training, validation, and test sets.

2. Problem identification

In this section you should discuss in detail and in your own words the problem identified in Figure 1 of the coursework specs. You also need to discuss “why” and “how” this problem occurs and “what” you can do to solve the problem. Following this discussion, you will report your baseline experiments for EMNIST by varying hidden unit sizes and the number of hidden layers. No need for theoretical explanations of things covered in the course, but if you go beyond what was covered please explain what you did with references to relevant paper(s) if appropriate.

In this section you should aim to cover both the “what” and the “why”: *what* you did, giving sufficient information (hyperparameter settings, etc.) so that someone else (e.g. another student on the course) could reproduce your results; and *why* you performed the experiments you are reporting - what you are aiming to discover what is the motivation for the particular experiments you undertook. You should also provide some discussion and interpretation of your results

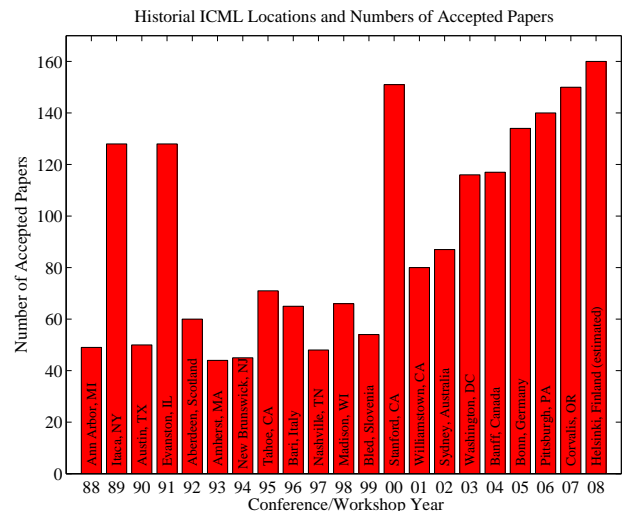


Figure 1. Historical locations and number of accepted papers for International Machine Learning Conferences (ICML 1993 – ICML 2008) and International Workshops on Machine Learning (ML 1988 – ML 1992). At the time this figure was produced, the number of accepted papers for ICML 2008 was unknown and instead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

and if these experiments worsen or mitigate the problem discussed above.

Your experimental sections should include graphs (for instance, figure 1) and/or tables (for instance, table 1)¹, using the `figure` and `table` environments, in which you use `\includegraphics` to include an image (pdf, png, or jpg formats). Please export graphs as **vector graphics** rather than **raster files** as this will make sure all detail in the plot is visible. Matplotlib supports saving high quality figures in a wide range of common image formats using the `savefig` function. **You should use `savefig` rather than copying the screen-resolution raster images outputted in the notebook.** An example of using `savefig` to save a figure as a PDF file (which can be included as graphics in a LaTeX document is given in the coursework document.

If you need a figure or table to stretch across two columns use the `figure*` or `table*` environment instead of the `figure` or `table` environment. Use the `subfigure` environment if you want to include multiple graphics in a single figure.

¹These examples were taken from the ICML template paper.

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	✓

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

3. Dropout and Weight Penalty

In this section you should analyze and discuss Dropout and Weight Penalty, introduce them either as equations or as algorithmic pseudocode. Discuss the benefits of both approaches, how they help in mitigating the problem and how do they compare with each other. If you present the different approaches as algorithms, you can use the `algorithm` and `algorithmic` environments to format pseudocode (for instance, Algorithm 1). These require the corresponding style files, `algorithm.sty` and `algorithmic.sty` which are supplied with this package.

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 $true$

You should, in your own words, explain what these functions do, and how they differ.

4. Balanced EMNIST Experiments

Here you should present your experiments and results, comparing and contrasting with different hyper-parameter settings and different combinations when applied to multi-layer neural networks. As before concentrate on the “what” (remember give enough information so someone can reproduce your experiments), the “why” (why did you choose the experiments that you performed – you may have been motivated by your earlier results, by the literature, or by a specific research question), and the interpretation of your results. In addition, you should provide enough details about your experiments (hyperparameters, network architecture, training details, etc) such that a reader can reproduce your results based on the description.

In every section, you should present your results in a way

that makes it easy for a reader to understand what they mean. You should facilitate comparisons either using graphs with multiple curves or (if appropriate, e.g. for accuracies) a results table. You need to avoid having too many figures, poorly labelled graphs, and graphs which should be comparable but which use different axis scales. A good presentation will enable the reader to compare trends in the same graph – each graph should summarise the results relating to a particular research (sub)question.

Your discussion should interpret the results, both in terms of summarising the outcomes of a particular experiment, and attempting to relate to the underlying models. A good report would have some analysis, resulting in an understanding of why particular results are observed and how they help in improving performance and overcoming the problem, perhaps with reference to the literature. Use `bibtex` to organise your references – in this case the references are in the file `example-refs.bib`. Here is an example reference (Langley, 2000).

5. Literature Review: Paper Title

In this section you should discuss one paper you chose from the list of 3 papers in the coursework specs. **Explain the paper in your own words.** The discussion should provide a summary of the paper, the strengths and limitations of the paper and what do you think can make it a better paper (if any missing experiments, visualization, theoretical proof etc.). Also remember to include the “what” and “why” of the experiments and a brief interpretation of the results, all in your own words.

6. Conclusions

Based on your experiments, what is the setup (models, algorithms) that produces your best system on the Balanced EMNIST data and helps to mitigate the problem?

You should draw conclusions from the experiments, relate to the research questions outlined in the introduction (section 1). You should state the conclusions clearly and concisely. It is good if the conclusion from one experiment influenced what you did in later experiments – your aim is to learn from your experiments. Extra credit if you relate your findings to what has been reported in the literature.

A good conclusions section would also include a further work discussion, building on work done so far, and referencing the literature where appropriate.

References

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.