MLP Coursework 2: Learning rules, BatchNorm, and ConvNets

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 2 report. This 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 should be no longer than **seven pages**, with an additional page 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. 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.

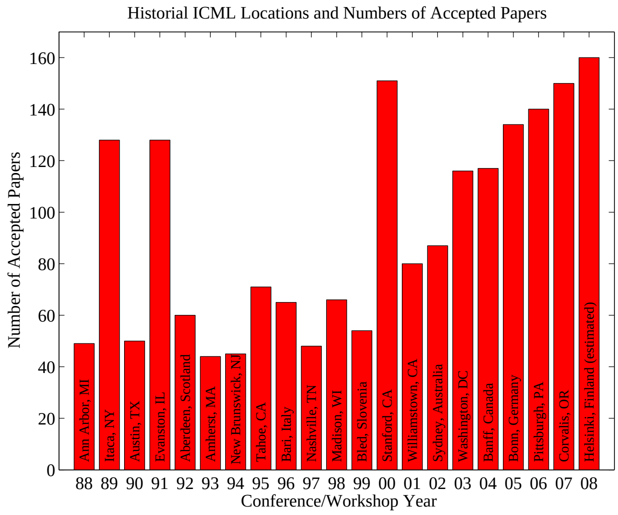

Figure 1. Historical locations and number of accepted papers for International Machine Learning Conferences (ICML 1993 – ICML 2008) and International Workshops on Machine Learning (ML 1988 – ML 1992). At the time this figure was produced, the number of accepted papers for ICML 2008 was unknown and instead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

2. Baseline systems

In this section you should report your baseline experiments for EMNIST. No need for theoretical explanations of things covered in the course, but should 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 the 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.

As before, 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 LaTeXdocument is given in the coursework 1 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.

3. Learning rules

In this section you should compare RMSProp and Adam with gradient descent, introducing these learning rules either as equations or as algorithmic pseudocode. 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.

You should, in your own words, explain what the differ-

[1]These 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.

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 $true$

ent learning rules do, and how they differ. You should then present your experiments and results, comparing and contrasting stochastic gradient descent, RMSProp, and Adam. 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 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, 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 a an example reference (Langley, 2000).

4. Batch normalisation

In this section you should present batch normalisation, supported using equations or algorithmic pseudocode. Following this present your experiments, again remembering to include the “what”, the “why”, and the interpretation of results.

5. Convolutional networks

In this section you should present your experiments with convolutional networks. Explain the idea of convolutional layers and pooling layers, and briefly explain how you did the implementation. There is no need to include chunks of code. You should report the experiments you have undertaken, again remembering to include *what* experiments you performed (include details of hyperparameters, etc.), *why* you performed them (what was the motivation for the experiments, what research questions are you exploring), and the interpretation and discussion of your results.

6. Test results

The results reported in the previous sections should be on the validation set. You should finally report results on the EMNIST test set using what you judge to the be the best deep neural network (without convolutional layers) and the best convolutional network. Again focus on what the experiments were (be precise), why you chose to do them (in particular, how did you choose the architectures/settings to use with the test set), and a discussion/interpretation of the results.

7. Conclusions

You should draw conclusions from the experiments, related 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.