
MLP Coursework 1: Activation Functions

s1754321

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. In particular, it structures the document into five sections (plus an abstract and the references) – you should keep to this structure for your report. 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 **six 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 – in this case comparison of the behaviour of the different activation functions, experimental investigation of the impact of the depth of the network with respect to accuracy, and experimental investigation of different approaches to weight initialisation. This section should also include a concise description of the MNIST task and data – be precise: for example state the size of the training and validation sets.

2. Activation functions

This section should cover the theoretical methodology – in this case you should present the four activation functions: ReLU, Leaky ReLU, ELU, and SELU. I didn't do it in this document, but the first time you use an acronym you should say what it stands for, for example Restricted Linear Unit (ReLU). You should use equations to concisely describe each activation function. For example, ReLU:

$$\text{relu}(x) = \max(0, x), \quad (1)$$

which has the gradient:

$$\frac{d}{dx} \text{relu}(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0. \end{cases} \quad (2)$$

The LaTeX for the derivatives is slightly more complicated. We provided definitions near the top of the file (the part before `\begin{document}`) for `\relu`, `\lrelu`, `\elu`, and

`\selu`. There is no need to discuss the unit tests for these activation functions in this report.

It is probably not needed in this report, but if you would like to include an algorithm in your report, please 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
```

3. Experimental comparison of activation functions

In this section you should present the results and discussion of your experiments comparing networks using the different activation functions on the MNIST task. As explained in the coursework document, you should use 2 hidden layers with 100 hidden units per layer for these experiments. You can compare the learning curves (error vs epoch) for training and/or validation, and the validation set accuracies.

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

¹These examples were taken from the ICML template paper.

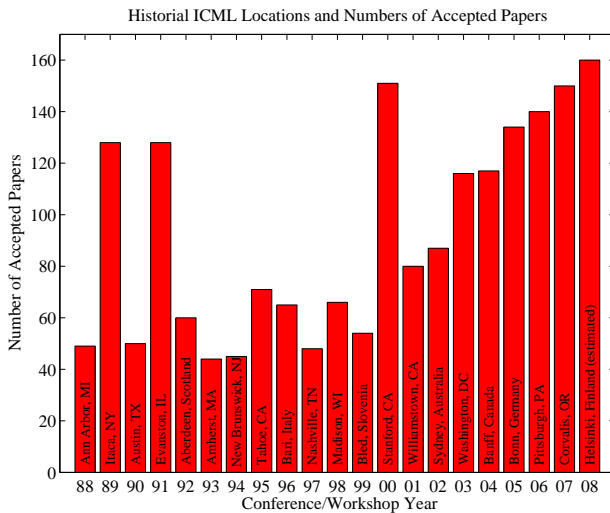


Figure 1. Historical locations and number of accepted papers for International Machine Learning Conferences (ICML 1993 – ICML 2008) and International Workshops on Machine Learning (ML 1988 – ML 1992). At the time this figure was produced, the number of accepted papers for ICML 2008 was unknown and instead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

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.

figure or table environment. Use the subfigure environment if you want to include multiple graphics in a single figure.

4. Deep neural network experiments

This section should report on your experiments on deeper networks for MNIST. The two sets of experiments are to explore the impact of the depth of the network (number of hidden layers), and a comparison of different approaches to weight initialisation.

In this section, and in the previous section, you should present your experimental results clearly and concisely, followed by an interpretation and discussion of results. You need to 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 an example reference (Langley, 2000).

5. 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.