Deep learning Assignment 2 (2020)

MPhil in Computational Biology

December 2, 2020

If there are errors found, I will update the assignment. **Due date: 2021-01-31 23:45**

Please submit your report to moodle, anonymised as before. Your report must be a maximum of ten pages, excluding the appendix. Your appendix should contain only a copy of your code.

This assignment is worth 50% of your overall mark for this module.

1 Hopfield network [25 marks]

Construct a Hopfield network with binary units, and test its ability to recall binary-valued input patterns. Comment on the following features of your model:

1. Storage capacity: how many patterns can it store? How does the sparseness (fraction of units set to +1 rather than -1) of the pattern affect this result?

2. Robustness: how is pattern recall affected by the random loss of weights?

3. **(Advanced:)** Explore an alternative method for setting the weights and see how it affects network performance.

Consult Chapter 42 of David Mackay's book as well as the course notes to get started.

2 Word2vec [15 marks]

Write a three page (maximum) summary of the word2vec algorithm.

In particular: describe the aim of the word2vec system. Run word2vec for yourself to investigate whether you can replicate key findings (such as the "X = France - Paris + Italy" example, or if words close to each other are semantically related). Find and summarise examples from the literature of how it can be used to solve interesting problems in computational biology.

You may use an existing R package like `https://github.com/bnosac/word2vec`. You do not need to train your own networks, but you can reuse the vectors that have already been created on large corpuses of data.

3 Essay [10 marks]

Write a two page essay on the challenges for reproducibility in the deep learning community. Include citations for your work. Some citations that you may wish to check include:

Lipton ZC, Steinhardt J (2019) Troubling Trends in Machine Learning Scholarship. Queueing Syst 17:45–77 Available at: `https://doi.org/10.1145/3317287.3328534`.

Hartley M, Olsson TSG (2020) dtoolAI: Reproducibility for Deep Learning. Patterns Prejudice 1:100073 Available at: `http://www.sciencedirect.com/science/article/pii/S2666389920300933`

Liu C, Gao C, Xia X, Lo D, Grundy J, Yang X (2020) On the Replicability and Reproducibility of Deep Learning in Software Engineering. arXiv [csSE] Available at: `http://arxiv.org/abs/2006.14244`.