Per-Sample Amnesiac Unlearning

July 14, 2024

Davide Marincione

Abstract

Typical machine unlearning techniques do not scale well with modern deep learning models, as these either require models to abide to strict theoretical guarantees (lowering their effectiveness) or be retrained (unfeasible for big models). In this project we propose a variation of Amnesiac Unlearning, a technique which promises to bring unlearning with little to no compromises to neural networks. Our method lowers the memory requirements of the original technique while maintaining its effectiveness. We test it on MNIST and CIFAR-100, showing that it can make a model forget samples and classes while maintaining its performance on the rest of the dataset.

1. Introduction

Machine Unlearning is an ever-more important topic in the field of Machine Learning. Not only because privacy concerns regarding deep learning techniques are increasing, but also because regulations such as the EU's GDPR oblige companies to delete user's data at their request. Because of this, research in trying to produce an unlearning procedure that requires the less possible amount of retraining is increasing.

As far as single-technique driven methods go (as combined procedures are often used in practice), the 'model offset' family of techniques is one of the most promising.

2. Project submission

Once you have finished your project, you are required to submit a report using this template. In principle, the only files you need to modify are "main.tex" and "bibliography.tex" (see Section 4 for more details). Hence, you can simply take this document and modify it directly with your own content.

Limits. The report must have at most 2 pages, without counting the bibliography, which can go to page 3. If you did your project with another student, then your budget is extended to at most 3 pages plus bibliography.

Code. Submitting your code is part of the exam. It need not be public (although we encourage it), but at least we must be able to access it for proper evaluation. One possibility is to put your code on a github repository. Please link the code directly from the report, example link: `https://github.com/erodola/DLAI-s2-2024`.

3. Report structure

There is no mandatory structure to adopt, but a typical report should look as follows. 1) An **introduction** section where the problem is presented, and the overall proposed approach is briefly described; 2) a **related work** section; 3) a **method** section, where the main methodology used for the project is described; 4) experimental **results** (qualitative, quantitative, or both); 5) **discussion and conclusions**.

4. Using LaTeX

If this is your first time using LaTeX, here are a few common instructions that you may find useful. Please read this while looking at the source code.

Formulas. You can write formulas inline, such as x^2, or you can put them in their own environment, for example:

$$\lambda_i = \int_{\mathcal{X}} \langle \nabla\phi_i(x), \nabla\phi_i(x) \rangle dx \,. \tag{1}$$

You can also refer to formulas without having to write their equation number by hand, such as Equation (1).

Figures. You can and are encouraged to include figures. See an example with Figure 1.

Table. Tables can be used to report quantitative results, here is one random example:

Bibliography. This is an example bibliographic reference (Anderson, 2008). If you want to add more, you must

Email: Davide Marincione <marincione.1927757@studenti.uniroma1.it>.

Deep Learning and Applied AI 2024, Sapienza University of Rome, 2nd semester a.y. 2023/2024.

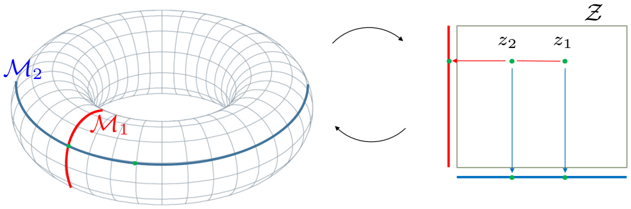

Figure 1. In the figure caption, you can write what you want including formulas, e.g. $\mathcal{X} \subset \mathbb{R}^3$. Notice that in this figure, we added mathematical symbols on top of the image by using the overpic command.

Table 1. Performance comparison.

#factors	β VAE	DCI Dis.	MIG	MIG-PCA	MIG-KM
One	100%	**99.0%**	63.7%	73.5%	69.2%
Variable	98.9%	94.9%	62.3%	70.5%	**66.9%**

edit the file “references.bib”.

References

Anderson, C. The end of theory: The data deluge makes the scientific method obsolete. *Wired magazine*, 16(7): 16–07, 2008.