The Stability Gap in Continual Learning with Deep Neural Networks

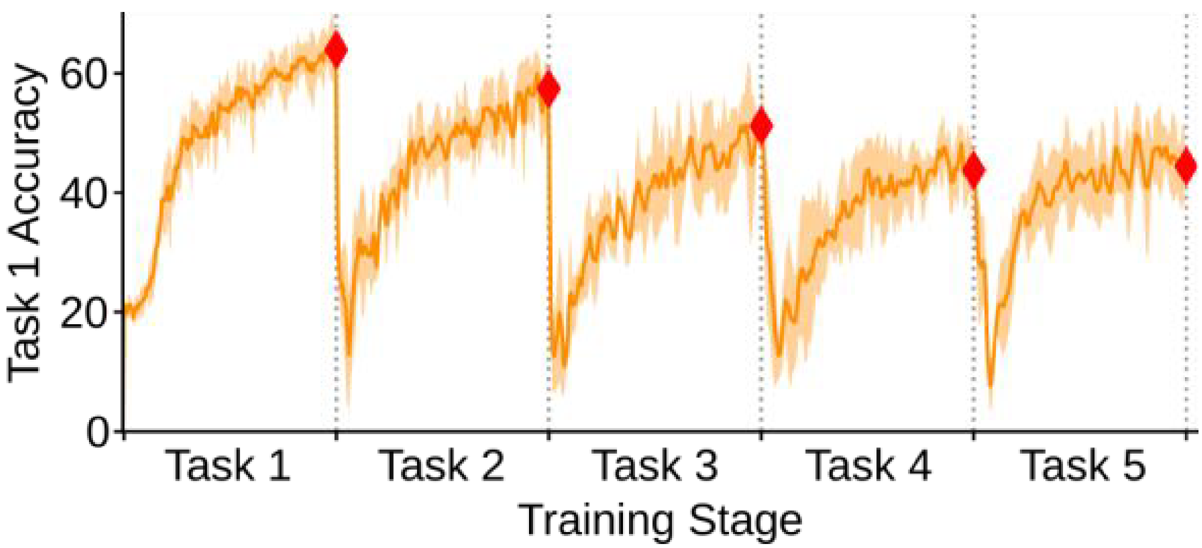
Responsible Professor: Tom Viering, t.j.viering@tudelft.nl

Supervisor: Gido van de Ven, G.M.vandeVen@tudelft.nl

Background and Motivation

Continually learning is a setting where neural networks learn a sequence of tasks. This is challenging for deep neural networks, because they forget past tasks when learning a new one. One of the main solutions is replay, which involves storing past tasks in a memory and revisiting them when training on later tasks. However, we recently discovered that replay still suffers from substantial forgetting, but that this forgetting is temporary. This phenomenon is called the 'stability gap' [1].

The stability gap is illustrated in the figure below [1]. Shown in this figure is the performance on the first task when a neural network is sequentially trained on the first five tasks, while using replay. The dashed vertical lines in this figure indicate task switches. Whenever the network starts to learn a new task, there is substantial but temporary forgetting of the first task.



The stability gap is problematic for safety-critical applications where sudden drops in performance could be harmful. If continual learning approaches are applied in, for example, medical applications, guarantees are needed with regards to their worst-case performance. Another reason we might want to avoid the stability gap relates to efficiency. It seems wasteful to first forget something, to then have to re-learn it later [2].

The goal of this research project is to gain deeper insight into the stability gap, and to lay the groundwork for developing new methods that can overcome the stability gap.

Research Questions for the Sub-Projects

Q1: How does the learning rate influence the stability gap? Can learning rate scheduling help to reduce or avoid the stability gap?
Q2: How does momentum and/or the type of optimizer (e.g., adam, adaGrad, SGD) influence the stability gap?
Q3: How does the mini-batch size influence the stability gap?
Q4: To what extent can the method Layerwise Proximal Replay (proposed in [3]) reduce the stability gap? (code is available on github)
Q5: To what extent can using second-order optimization methods help to reduce the stability gap? (Hint: one option could be to use the old task's Fisher Information matrix as a preconditioner when training on the next task, similar as done in [4]. In [4] a KFAC-approximation of the Fisher is used, but for this project using a diagonal approximation would be fine.)
Q6 (bonus): Can you propose and test your own method to reduce the stability gap?

All questions focus on different aspects that might influence the stability gap, but there is much collaboration possible. In particular, during the first part of the project the students can collaborate when setting up a "baseline" continual learning experiment in which there is a clear stability gap. Using that experiment as a starting point, each student should then run their own experiments to address their specific question.

Prerequisites

Enthusiastic to build careful experiments and learn to do empirical research.

Students will learn to implement continual learning experiments with deep neural networks in PyTorch. Prior experience with PyTorch is not necessary, we will teach you its use during the research project. The code for this does not need to be developed from scratch; code available from GitHub (e.g., https://github.com/GMvandeVen/continual-learning) can be used as a starting point.

Publicly available datasets (e.g., MNIST or CIFAR) will be used.

Q&A Sessions

March 13, 11:20-11:40 via Zoom:
https://tudelft.zoom.us/j/99866270436?pwd=yqODzr5HnzFrNULgAg81baq7m62OWV.1

References

[1] De Lange M, van de Ven GM, Tuytelaars T (2023), "Continual evaluation for lifelong learning: Identifying the stability gap", International Conference on Learning Representations (ICLR).
PDF: https://openreview.net/pdf?id=Zy350cRstc6

[2] Hess T, Tuytelaars T, van de Ven GM (2023), “Two complementary perspectives to continual learning: ask not only what to optimize, but also how”, Proceedings of the 1st ContinualAI Unconference, PMLR 249: 37-61.
PDF: https://arxiv.org/pdf/2311.04898

[3] Yoo J, Liu Y, Wood F, Pleiss G (2024), “Layerwise Proximal Replay: A Proximal Point Method for Online Continual Learning”, International Conference on Machine Learning (ICML).
PDF: https://arxiv.org/pdf/2402.09542

[4] Kao TC, Jensen K, van de Ven G, Bernacchia A, Hennequin G (2021), “Natural continual learning: success is a journey, not (just) a destination”, Advances in Neural Information Processing Systems, 34: 28067-28079.
PDF: https://arxiv.org/pdf/2106.08085