Analysis of Class Incremental Learning: Role of Learning Rate, Number of Classes, and Dataset Size

Aditya Somasundaram and Sai Vignesh

Columbia University

April 29, 2025

Motivation: Why Class-Incremental Learning?

- Class-Incremental Learning (CIL) addresses a critical challenge: how to learn new concepts while retaining knowledge.
- An evolving and ever-learning system is the dream
- Mirrors how humans learn throughout their lives

The Fundamental Challenge: Catastrophic Forgetting

Learning new classes makes the model forget the old ones.

A simple illustration of this:

- When a neural network trained on task A is subsequently trained on task B
- Performance on task B improves rapidly
- Performance on task A declines precipitously
- The model becomes specialized to the most recent task
- Previously acquired knowledge is largely lost

The Fundamental Challenge: Catastrophic Forgetting

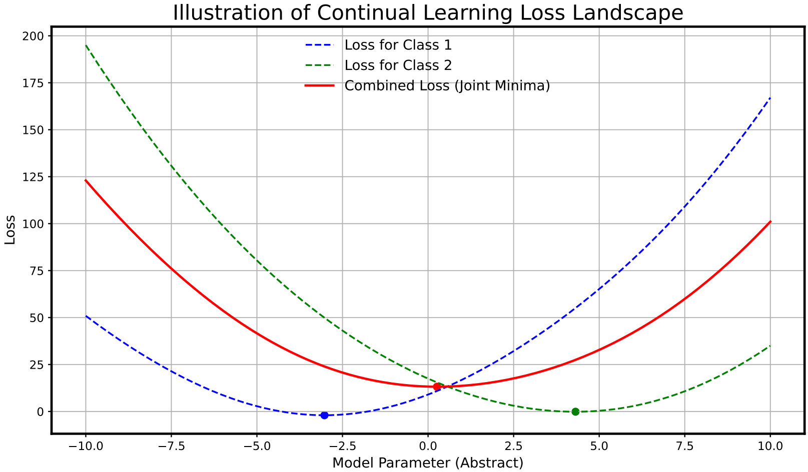


Figure 1: A toy example of the loss landscapes. It is difficult to bounce between the two individual curves and get to the joint minima.

The Fundamental Challenge: Catastrophic Forgetting

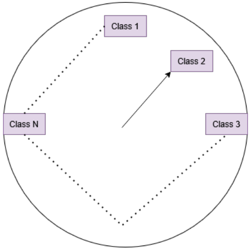
Why does a model forget?

Reasons

- Neural networks optimizing for current task data
- Limited data representation of old classes during new training
- Parameter updates that overwrite crucial information

Our analysis: How does Forgetting Evolve?

We perform our analysis by training classes in a cyclic fashion.



Interesting questions we aim to answer

- What about having a large sample size per class?
- What about the learning rate?
- Increasing the number of classes definitely increases the complexity in learning (makes sense intuitively). But by how much?

The setup

Occam's Razor: simplicity is key

- We select the recently introduced MNIST-1D dataset: this has 40 dimensional vectors (reduced from MNIST's 784)

Our model has fixed size: $40 \to 128 \to 10$, and fixed batch size of 32. We train for a total of 5000 cycles. We perform experiments systematically by varying 3 key factors:

- Total number of classes $\{2, \cdots, 10\}$
- Total samples per class per cycle $\{2^5, 2^6, 2^7, 2^8, 2^9, 2^{10}\}$
- Learning rates $\{10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$

Initial Experiments: Overall Accuracy

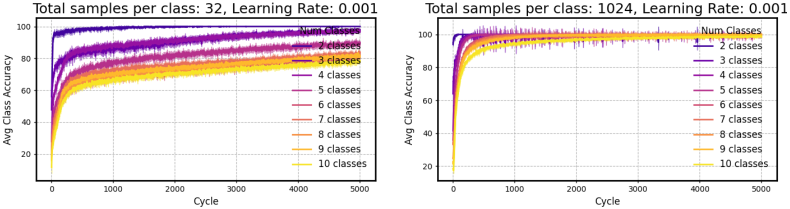


Figure 2: Accuracy versus training cycle is shown above for (sample size per class, learning rate) pairs of (32, 10^{-3}) and (1024, 10^{-3}). Experiments with fewer total classes converge faster. If the total samples per class per cycle is higher, the learning process if overall faster.

Initial Experiments: Class-wise Accuracy

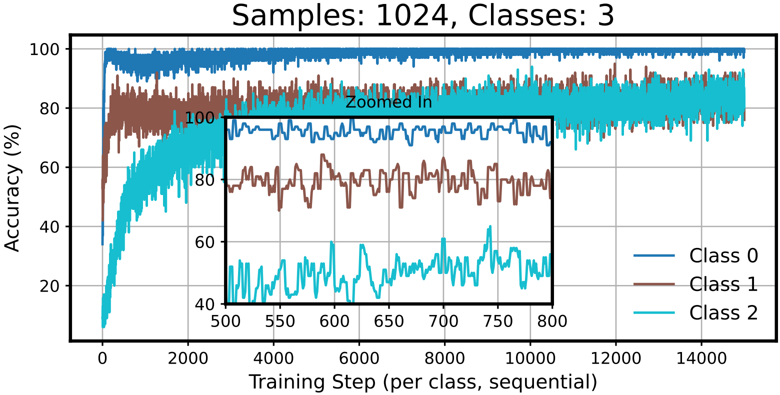


Figure 3: Zoomed in plot of class wise accuracy noted after training on each class in the cyclic training scheme. Note how the accuracies are very noisy with training, indicating repeated forgetting and relearning of data.

Evolution of Forgetting: Decreases, but does not reach zero!

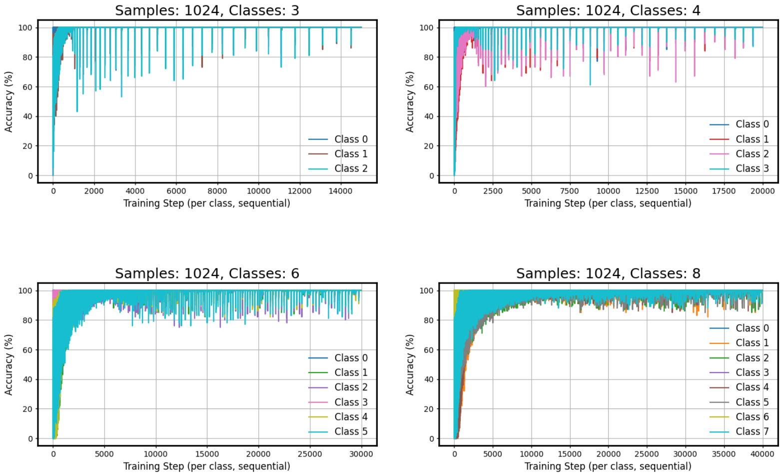


Figure 4: Forgetting analysis on 1024 samples and learning rate of 10^{-3} while varying number of classes. Note how after a series of training steps, there is a constant non zero forgetting!

Latching: First Impressions Last

Figure 5: Observe how the model learns the first class it sees and "latches" onto it. This behavior is observed in most of our experiments (recall Figure 3?)

Comparison to Baseline

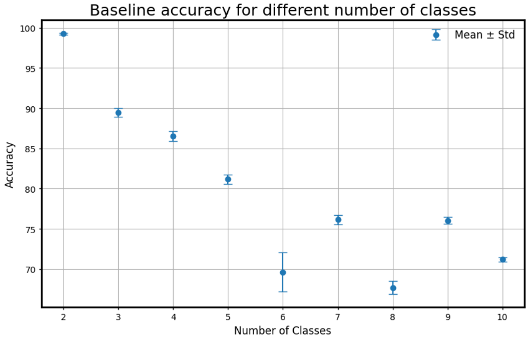


Figure 6: It is interesting to note that the authors are also only able to get 68% using an MLP on this dataset (with 10 classes). Surprisingly, some configurations of the continuous learning setup can beat this. [1]

[1] Our testing mechanism is pulls 100 samples per class at random. It is possible that evaluating the entire dataset will not yield such a result. We plan to do this before reporting this anomaly with confidence.

Conclusion and Future Work

- Analysis of strange phenomena: latching behavior, last-in-first-out forgetting pattern
- Possible scaling laws (total number of samples per class, number of classes, learning rate)
- Move into Task Incremental Learning from Class Incremental Learning (combine classes into tasks)

Thoughts/Questions?

Thank You!