



UNC  
NLP

# Towards Scalable Exact Machine Unlearning Using Parameter-Efficient Fine-Tuning

Somnath Basu Roy Chowdhury, Krzysztof Choromanski, Arijit Sehanobish, Avinava Dubey, and Snigdha Chaturvedi  
{somnath, snigdha}@cs.unc.edu

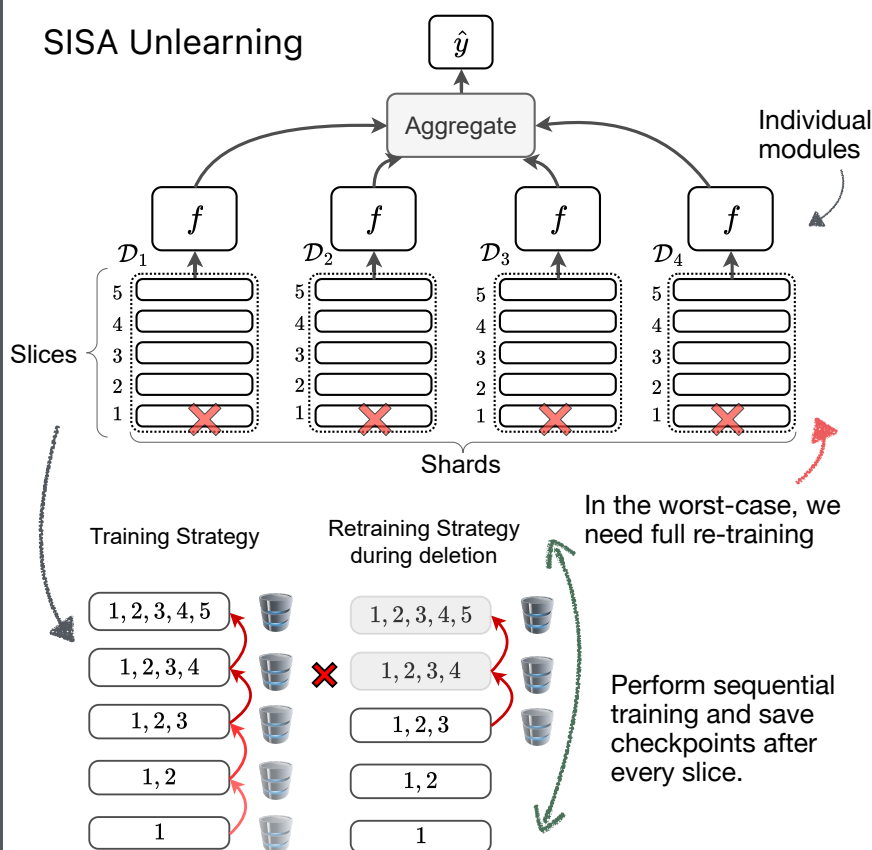


## Machine Unlearning

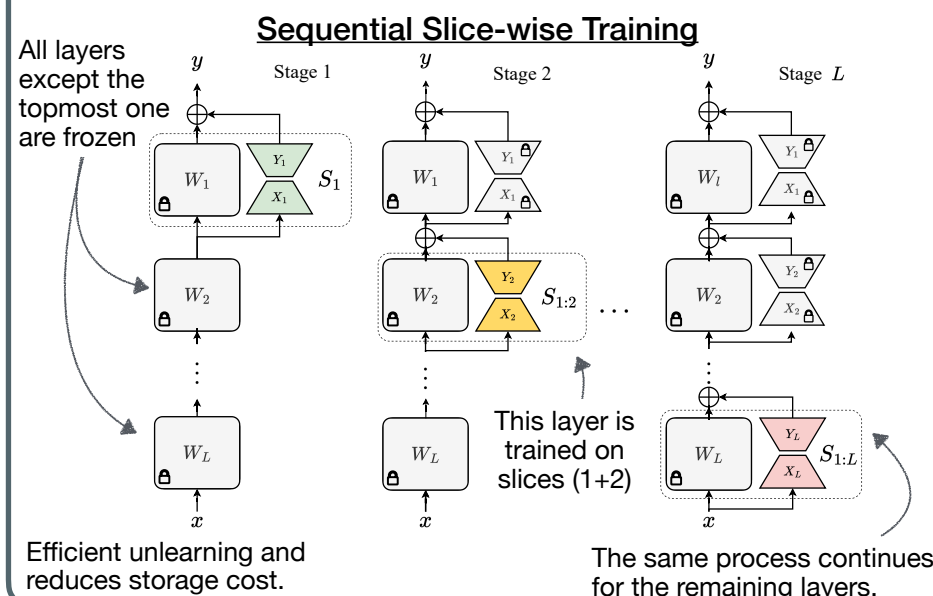
- What is machine unlearning?
  - Machine unlearning involves removing the influence a training data instance from a trained model
- What is exact unlearning?
  - Exact unlearning guarantees removal of the data instance by re-training model components
  - Performed by using a modular machine learning system

## Exact Unlearning Techniques

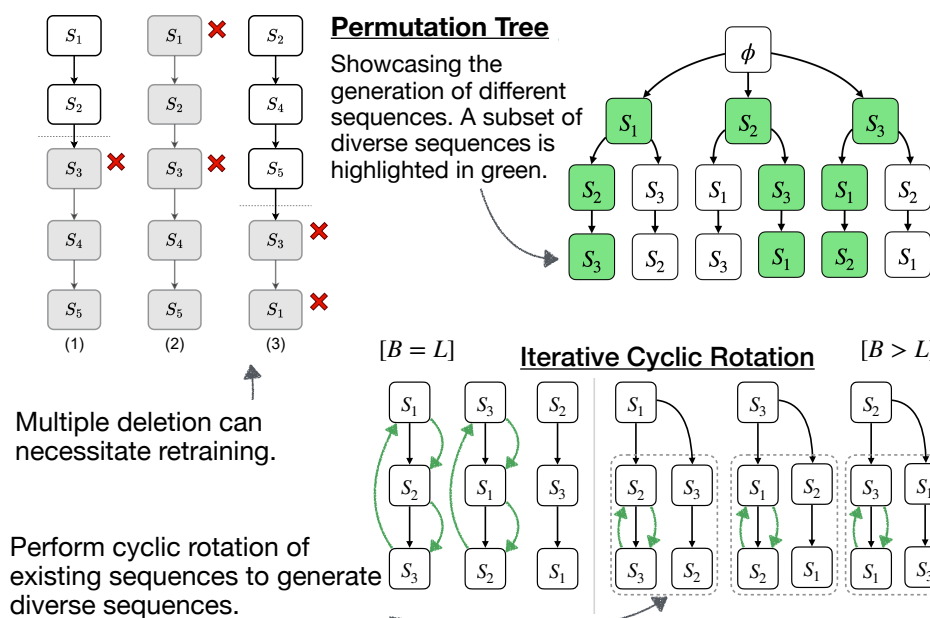
### SISA Unlearning



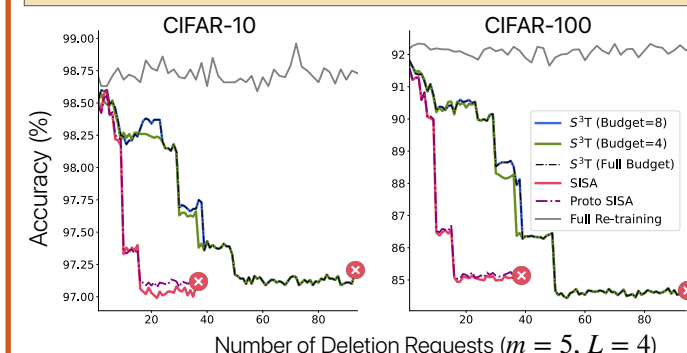
## S<sup>3</sup>T: Sequence-aware Sharded Sliced Training



## Training Sequence Selection



## Empirical Results



(a) Plot indicates how performance degrades as more deletion requests are processed. (b) S<sup>3</sup>T achieves better performance and process more deletion requests compared to baselines.

## Theoretical Guarantees

- Deletion rate of S<sup>3</sup>T:  $O(mL \log(mL \min(B, L)))$   
To achieve the best deletion rate we do not need a budget more than  $L$
- Deletion rate of SISA:  $O(mL \log(mL))$
- Performance gap between S<sup>3</sup>T and SISA:  $\zeta(1 - \zeta^{B'})$ , where  $\zeta$  is a positive fraction and  $B' \propto B$  (training budget)

## Conclusion

- We present a practical way to convert ML systems into modular ones
- We achieve modularity by using:
  - PEFT components
  - Sequential training for parameter isolation
  - Using diverse training sequences to reduce re-training cost



[Link to Paper!](#)