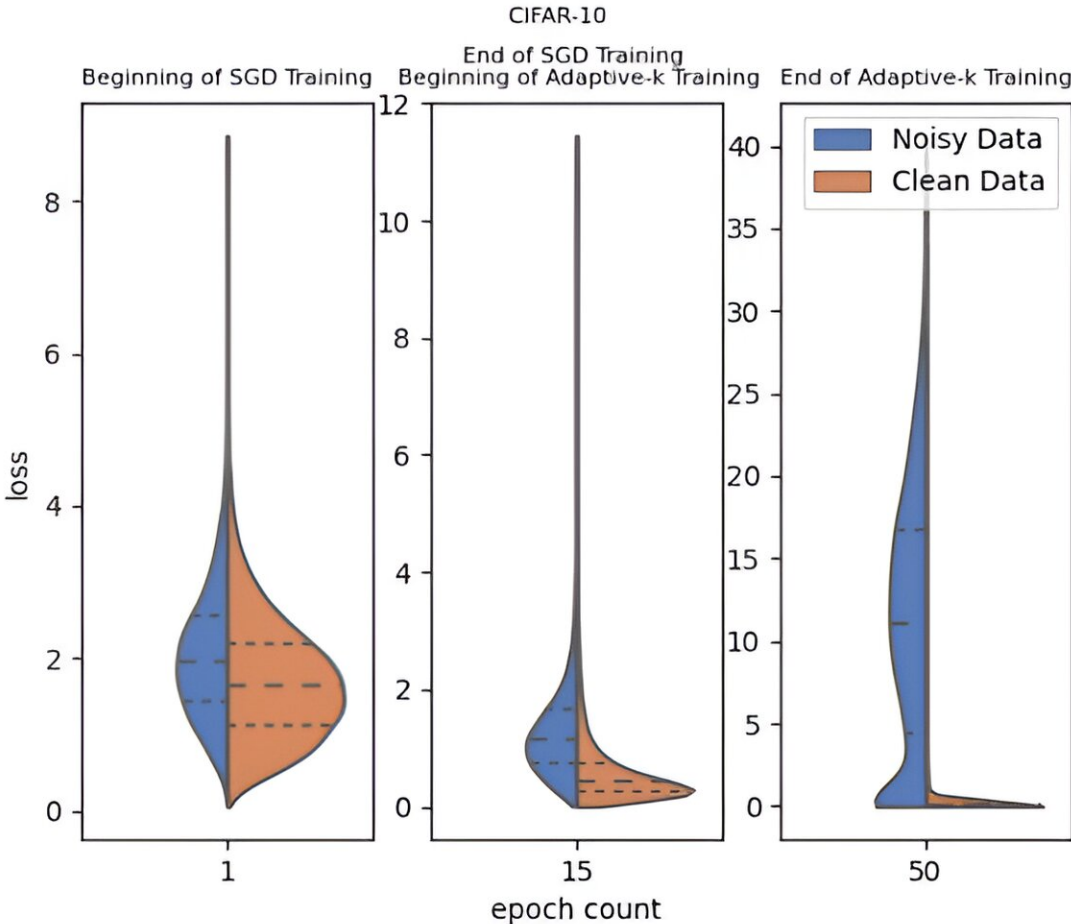# Adaptive-k: A simple and effective method for robust training in label noisy datasets

August 22 2024

Loss Distributions of Clean and Noisy Data Throughout Adaptive-k Training for Noise Ratio τ = 0.3. Credit: Enes Dedeoglu, Himmet Toprak Kesgin, Mehmet Fatih Amasyali

Training deep learning models on large datasets is essential for their success; however, these datasets often contain label noise, which can significantly decrease the classification performance on test datasets.

To address this issue, a research team consisting of Enes Dedeoglu, H. Toprak Kesgin, and Prof. Dr. M. Fatih Amasyali from Yildiz Technical University developed a method called Adaptive-k, which improves the optimization process and yields better results in the presence of label noise.

Their research was published in *Frontiers of Computer Science*.

The Adaptive-k method stands out by adaptively determining the number of samples selected for updating from the mini-batch, leading to a more effective separation of noisy samples and ultimately increasing the success of training in label noisy datasets.

This innovative method is simple, effective, and does not require prior knowledge of the dataset's noise ratio, additional model training, or significant increases in training time. Adaptive-k has demonstrated its potential to revolutionize the way [deep learning models](#) are trained on noisy datasets by showing performance closest to the Oracle method, where noisy samples are entirely removed from the dataset.

In their research, the team compared the Adaptive-k method with other popular algorithms, such as Vanilla, MKL, Vanilla-MKL, and Trimloss, and assessed its performance in relation to the Oracle scenario, where all noisy samples are known and excluded.

Experiments were conducted on three image datasets and four text datasets, proving that Adaptive-k consistently performs better in label noisy datasets. Furthermore, the Adaptive-k method is compatible with various optimizers, such as SGD, SGDM, and Adam.

Sample Ratio Selected as Clean by Adaptive-k Throughout Epochs
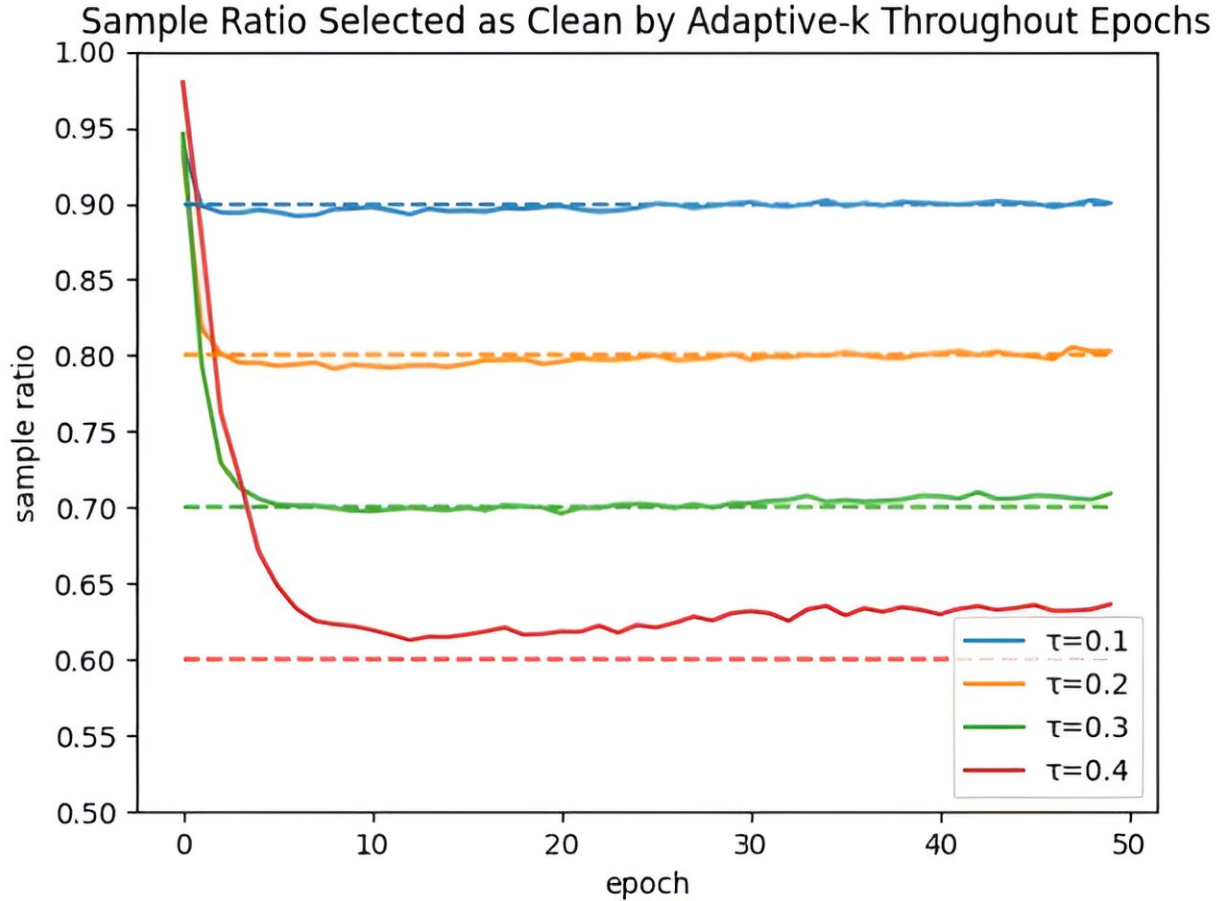
demonstrates how Adaptive-k accurately estimates clean sample ratios during adaptive training for the MNIST dataset. Credit: Enes Dedeoglu, Himmet Toprak Kesgin, Mehmet Fatih Amasyali

The primary contributions of this research include:

- Introducing Adaptive-k, a novel algorithm for robust training of label noisy datasets, which is easy to implement and does not require additional model training or data augmentation.
- Theoretical analysis of Adaptive-k and comparison with the MKL algorithm and SGD. • High accuracy noise ratio estimation

using Adaptive-k without prior knowledge of the dataset or hyperparameter adjustments.

- Empirical comparisons of Adaptive-k with Oracle, Vanilla, MKL, Vanilla-MKL, and Trimloss algorithms on multiple image and text datasets.

Future research will focus on refining the Adaptive-k method, exploring additional applications, and further enhancing its performance.

More information: Enes Dedeoglu et al, A robust optimization method for label noisy datasets based on adaptive threshold: Adaptive-k, *Frontiers of Computer Science* (2023). DOI: 10.1007/s11704-023-2430-4

Provided by Higher Education Press