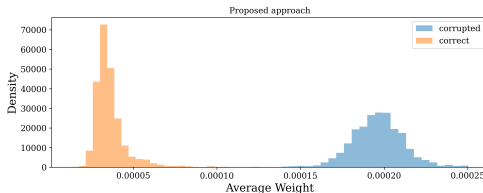
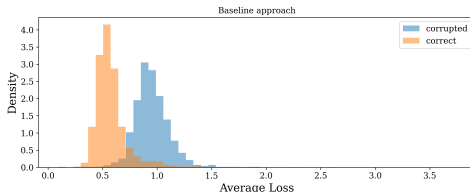


# Identifying Corrupted Labels using Loss Values

Dataset:  $(X, \mathbf{y}) = \{(x_i, y_i)\}_{i=1}^n$ , where some labels  $y_i$  may be corrupted.

**Hypothesis:** Incorrectly labeled samples  $(x_i, y_i)$  often yield higher loss values  $\mathcal{L}(f(x_i), y_i)$  during model training.

**Method:** Increase the loss gap between correct and incorrect labels using the ALSO optimizer<sup>a</sup>.



---

<sup>a</sup>A. Beznosikov et al. Mirror-Prox

Algorithm with Linear Convergence Rate and its Application for Dynamic Loss Scaling // 2025.