Robust Binary Loss for Multi-category Classification with Label Noise

Defu Liu[1], *Guowu Yang*[1], *Jinzhao Wu*[2], *Jiayi Zhao*[3], Fengmao Lv[3*]

[1] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, China

[2] School of Mathematics and Physics, Guangxi University for Nationalities, Nanning, 530004, China

[3]Center of Statistical Research and School of Statistics, Southwestern University of Finance and Economics, Chengdu, 611130, China

gxdefu@gmail.com, guowu@uestc.edu.cn, gxmdwjzh@aliyun.com, zjy3587712@163.com, fengmaolv@126.com

Paper Number: 4851

ICASSP 2021 TORONTO Canada June 6-11, 2021 Metro Toronto Convention Centre

Background

Thought deep model achieved tremendous success, two challenges remain for traditional deep learning:

- The generalization performance of deep model heavily depend on large-scale accurately labeled data.
- The training data often suffer from label noise in many applications.

Hence, it is essential to explore methods that can train deep models effectively under label noise.

Motivation

- For the case of binary classification, it has been shown that binary symmetric loss function can be noise-tolerant.
- Motivated by this observation, we wonder whether the robustness of the noise-tolerant binary loss functions can be generalized to the case of multi-category classification.

Our approach

- This paper proposes to tackle K-class classification problems by employing K binary classifiers.
- Each binary classifier $g_k(x)$can be regarded as a scoring function that reveals how likely a sample belongs to category k versus the rest $K-1$ categories.
- Then, the multi-category classifier $f(x)$ have the following formulation: $f(\boldsymbol{x}) = \underset{k\in\{1,\cdots,K\}}{\arg\max}\ g_k(\boldsymbol{x})$
- Finally, We leverage the multi-category large margin classification approaches, i.e. Pairwise-Comparison (PC) or One-Versus-All (OVA), to jointly train the binary classifiers for multi-category classification.
- Together with binary symmetric loss function, the objective loss can be formulated as follows:

$$\mathcal{L}_{\mathrm{OVA}}(f(\boldsymbol{x}), y_{\boldsymbol{x}}) = \frac{1}{K-1}\sum_{y'\neq y_{\boldsymbol{x}}} \ell\left(-g_{y'}(\boldsymbol{x}), 1\right) + \ell\left(g_{y_{\boldsymbol{x}}}(\boldsymbol{x}), 1\right),$$

$$\mathcal{L}_{\mathrm{PC}}(f(\boldsymbol{x}), y_{\boldsymbol{x}}) = \sum_{y'\neq y_{\boldsymbol{x}}} \ell\left(g_{y_{\boldsymbol{x}}}(\boldsymbol{x}) - g_{y'}(\boldsymbol{x}), 1\right), \ell \in \{\ell_{\mathrm{sig}}, \ell_{\mathrm{ramp}}\}.$$

Experimental results

- **Dataset & label noise.** We use MNIST, FASHION-MNIST, and CIFAR-10 to verify our approach. We verify the robustness of our approach against the symmetric noise, the simple non-uniform noise.

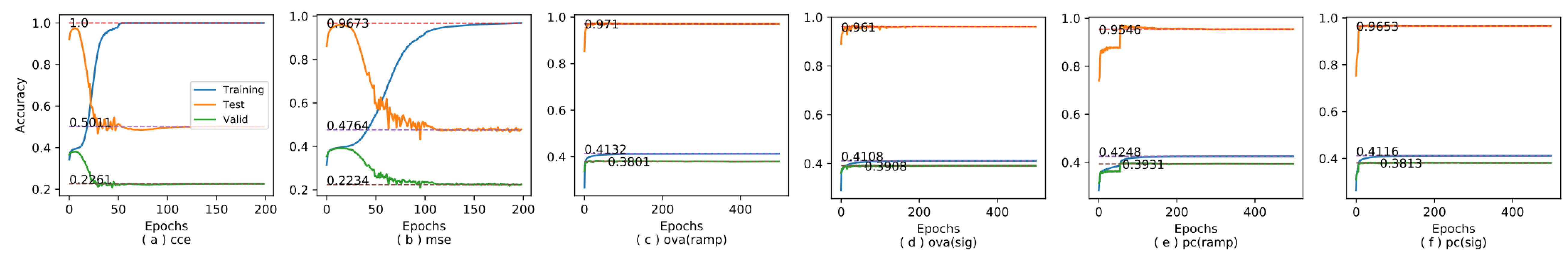

Fig. 1. Classification accuracy vs training epochs over the MNIST dataset with $\eta = 60\%$ symmetric noise.

- **The results clearly verify that our proposed loss functions are noise-tolerant.**

Label noise

Comparison methods

We compare the performance of our approach with different loss functions, including the CCE, MSE, and MAE loss.

- CCE and MSE are standard loss functions widely used in machine learning.
- The MAE loss has been shown to be robust against label noise in multicategory classification problems.

Main reference

[1] Ghosh et al., "Robust Loss Functions under Label Noise for Deep Neural Networks", AAAI, 2017.

[2] Ghosh et al., "Making Risk Minimization Tolerant to Label Noise", Neurocomputing, 2015.