A Study of Learning with Noisy Labels

Jianmei Ye, Master of Computer Science, Fulton School of Engineering, ASU

Abstract

The primary objective to this research was to study and understand the binary classification problem in the presence of random class-conditional noise. More specifically, in the learning of such random labels that have independently been flipped with some small probability.
The state-of-art algorithm from the paper "Learning with Noisy Labels" " (abbreviated LNL) , is the first one providing guarantees for risk minimization under random label noise without any assumption on the true distribution.
Index Terms — Binary Classification, Class-Conditional, Empirical Risk Minimization, Random Noise

Introduction

The goal for the methodology is to deal with situation when the binary classification training label are not reliable and to ensure that the average risk is minimized.
The learning algorithms will draw iid samples from a noisy data D_ρ of a "clean" distribution $\mathfrak{D}$, where the noise rate ρ ($\boldsymbol{\rho_{+1}} + \boldsymbol{\rho_{-1}} < 1$) depends on the class label. And it will tune the noise rate parameter though cross-validation.

Method

Lemma 1.

Let $\ell(\mathrm{t}, \mathrm{y})$ *be any bounded loss function. Then, if we define,*

$$\tilde{\boldsymbol{\ell}}(\mathbf{t}, \mathbf{y}) := \frac{(\mathbf{1} - \boldsymbol{\rho}_{-\mathbf{y}})\boldsymbol{\ell}(\boldsymbol{t}, \boldsymbol{y}) - \boldsymbol{\rho}_{\boldsymbol{y}}\boldsymbol{\ell}(\boldsymbol{t}, -\boldsymbol{y})}{\mathbf{1} - \boldsymbol{\rho}_{+\mathbf{1}} - \boldsymbol{\rho}_{-\mathbf{1}}}$$

we have, for any t, y, $\mathbb{E}_{\tilde{y}}\left[\tilde{\ell}(\mathrm{t}, \tilde{\mathrm{y}})\right] = \ell(\mathrm{t}, \mathrm{y})$

where

$\tilde{\ell}(\mathrm{t}, \tilde{\mathrm{y}})$ *denote a suitably modified* ℓ for use with noisy lables ρ_{+1} *and* ρ_{-1} *are known nosie rate,* $\tilde{y}$ *denotes noisy labels*

A good predictor in the presence of label noise can be learned by minimizing the sample average

$$\tilde{f} \leftarrow \underset{f \in \mathcal{F}}{argmin}\, \hat{R}_{\tilde{L}}(f) \qquad (*)$$

Thus, we then can know that for any fixed $\boldsymbol{f} \in \boldsymbol{\mathcal{F}}$, the (*) sample average converges to $\boldsymbol{R\ell,D}(\boldsymbol{f})$ the $\boldsymbol{\ell}$ **– risk** under the Noise – free distribution $\mathfrak{D}$ by the unbiased $\boldsymbol{\ell}$ obtained from Lemma 1.

Implementation & Profiling

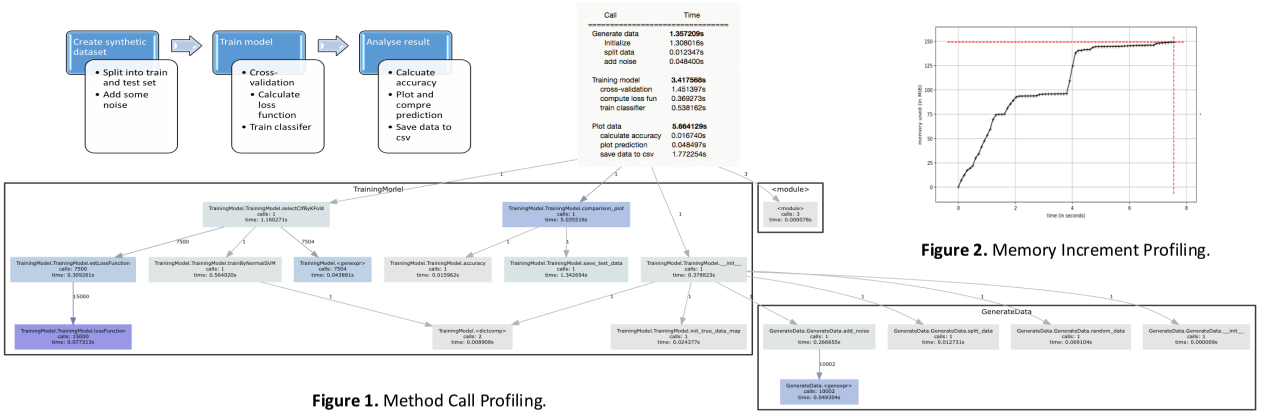

Figure 1. Method Call Profiling.

Figure 2. Memory Increment Profiling.

Experiment

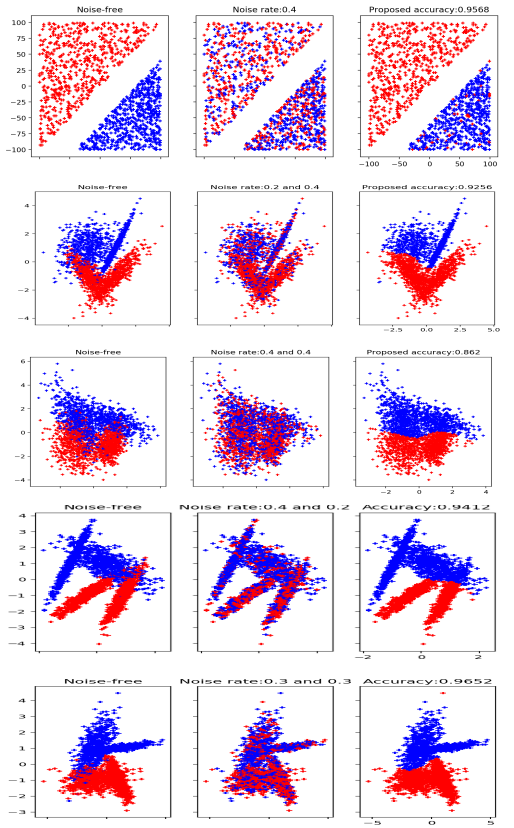

Figure 3. Noise Free VS. Noised Data VS. Predicted

Conclusions

This research is a comprehensive study and implementation of the proposed method I in LNL. The proposed algorithms addressed the effectiveness of risk minimization and it preforms competitive outcome at the situation when the random classification noise is present.