Learning with Bounded Instance- and Label-Dependent Label Noise
Supplementary Material

A. Proofs

A.1. Proof of Theorem 1

Proof. \( \forall x \in \text{supp}(P_D^\ast(X)) = \text{supp}(P_D(X)) \), we have
\[
g_D^\ast(x) = \text{sgn} \left( P_D^\ast(Y = +1|x) - \frac{1}{2} \right) = \text{sgn} \left( 1[g_D^\ast(x) = +1] - \frac{1}{2} \right) = g_D^\ast(x),
\]
where the last equality is justified by checking the possible binary values of \( g_D^\ast(x) \), e.g., when \( g_D^\ast(x) = +1 \), \( \text{sgn}(1[g_D^\ast(x) = +1] - \frac{1}{2}) = +1 \); when \( g_D^\ast(x) = -1 \), \( \text{sgn}(1[g_D^\ast(x) = +1] - \frac{1}{2}) = -1 \). \( \square \)

A.2. Proof of Theorem 2 and Corollary 1

Proof. \( \forall x \in X \), \( \tilde{\eta}(x) \) can be rewritten as
\[
\tilde{\eta}(x) = P(\tilde{Y} = +1, Y = +1|x) + P(\tilde{Y} = +1, Y = -1|x) = P(\tilde{Y} = +1|Y = +1, x)P(Y = +1|x) + P(\tilde{Y} = +1|Y = -1, x)P(Y = -1|x) = (1 - \rho_+(x))\eta(x) + \rho_-(x)(1 - \eta(x))
\]
Then, we have
\[
\eta(x) \geq \frac{1}{2} \implies \tilde{\eta}(x) = (1 - \rho_+(x))\eta(x) + \rho_-(x)(1 - \eta(x)) \geq (1 - \rho_+(x))\eta(x) \geq \frac{1 - UB(\rho_+(x))}{2}
\]
and its contrapositive
\[
\tilde{\eta}(x) < \frac{1 - UB(\rho_+(x))}{2} \implies \eta(x) < \frac{1}{2} \implies g_D^\ast(x) = -1
\]
The last step follows by Lemma 1. Similarly, we can prove \( \tilde{\eta}(x) > \frac{1 + UB(\rho_-(x))}{2} \implies g_D^\ast(x) = +1 \).

Corollary 1 holds by replacing \( UB(\rho_+(x)) \) and \( UB(\rho_-(x)) \) by \( \rho_{+1\text{max}} \) and \( \rho_{-1\text{max}} \), respectively. \( \square \)

A.3. Proof of Propositions 1 and 2

A.3.1. Proposition 1

Proof. The following Lemma holds because of the basic Rademacher bound on the maximal deviation between the expected and empirical risks (Bartlett & Mendelson, 2002).

Lemma A1. For any \( \delta > 0 \), with probability at least \( 1 - \delta \), we have
\[
\sup_{f \in F} \left| \tilde{R}_{D^\ast,L}^\ast(f) - R_{D^\ast,L}(f) \right| \leq R(L \circ F) + b\sqrt{\frac{\log(1/\delta)}{2m}}.
\]

Then, with probability at least \( 1 - \delta \), we have
\[
R_{D^\ast,L}(f) - R_{D^\ast,L}(\hat{f}_{D^\ast,L}) = (R_{D^\ast,L}(f) - \tilde{R}_{D^\ast,L}(f)) + (\tilde{R}_{D^\ast,L}(f) - R_{D^\ast,L}(f)) \leq 2\sup_{f \in F} \left| \tilde{R}_{D^\ast,L}(f) - R_{D^\ast,L}(f) \right| \leq 2R(L \circ F) + 2b\sqrt{\frac{\log(1/\delta)}{2m}},
\]
where the first inequality holds because \( \tilde{f}_{D^\ast,L} = \arg \min_{f \in F} R_{D^\ast,L}(f) \) and the second inequality follows by Lemma A1. \( \square \)

A.3.2. Proposition 2

Proof. Notice that \( R_{D,L}(f) = R_{D^\ast,L}(f) \), then the proof is similar with the proof of Proposition 1. \( \square \)

A.4. Proof of Theorem 3

Proof. \( \forall x \in X \), we have
\[
\tilde{\eta}(x) = (1 - \rho_+(x))\eta(x) + (1 - \eta(x))\rho_-(x) = (1 - \rho_+(x) - \rho_-(x))\eta(x) + \rho_-(x) \geq \rho_-(x),
\]
where the first equality has been derived in the proof of Theorem 2 and the inequality follows by our bounded total noise assumption \( 0 \leq \rho_+(x) + \rho_-(x) < 1 \). Similarly, we can prove \( \rho_+(x) \leq 1 - \tilde{\eta}(x) \). \( \square \)
B. Extension to the Multiclass Classification

By the one-vs.-all strategy, our algorithm can be easily adapted for multiclass classification. In the multi-class case, our Theorem 1 still holds and keeps the idea of learning with distilled examples justified. An example \((x, y)\) is distilled if \(g^*_D(x) = y\), where \(g^*_D(x) = \arg \max_i P_D(Y = i|x)\) is the Bayes optimal classifier under \(D\). Like in the binary case, ILN can be modeled by flip rates \(\rho_y(x) = P(\tilde{Y} \neq y|x, Y = y)\) and \(\rho_{-y}(x) = P(\tilde{Y} = y|x, Y \neq y)\). Let \(\eta_y(x) = P(Y = y|x)\) and \(\tilde{\eta}_y(x) = P(\tilde{Y} = y|x)\). Easy to derive that \(\tilde{\eta}_y(x) > \frac{1 + UB(\rho_{-y}(x))}{2} \implies \eta_y(x) > \frac{1}{2} \implies (x, y)\) is distilled. Hence, distilled examples can be collected out of noisy examples by thresholding \(\tilde{\eta}_y(x)\). Other parts of our algorithm can be performed without special adaptations.

References