Electronic Journal of Statistics
Vol. 12 (2018) 1779–1781
ISSN: 1935-7524
https://doi.org/10.1214/18-EJS1422

Corrigendum to "Classification with asymmetric label noise: Consistency and maximal denoising"*

Gilles Blanchard

Institut für Mathematik
Universität Potsdam
e-mail: gilles.blanchard@math.uni-potsdam.de

Clayton Scott

Electrical and Computer Engineering
University of Michigan, Ann Arbor
e-mail: clayscot@umich.edu

Abstract: We point out a flaw in Lemma 15 of [1]. We also indicate how the main results of that section are still valid using a modified argument.

Received March 2018.

The proof of Lemma 15 in [1] is invalid because it uses that the distribution of Y given $\tilde{Y}$ is independent of X. Unfortunately, this is not true even if $\tilde{Y}$ given Y is independent of X, as can easily be verified using Bayes rule.

Fortunately, the other results [1] that depend on Lemma 15 can still be proved, using an analogue to Lemma 15 for the label flipping model [2, 3]. We state this alternate lemma and offer a concise proof, using the notation of [1].

Lemma 1. *Let $(X, Y, \tilde{Y})$ be jointly distributed. Assume $\tilde{Y}$ given Y is independent of X, denote $\rho_i = \Pr(\tilde{Y} \neq i | Y = i)$ and assume $\rho_0 + \rho_1 < 1$. Then for any f,*

$$(1 - \rho_0 - \rho_1)(R_P(f) - R_P^*) = R_{\tilde{P},\alpha}(f) - R_{\tilde{P},\alpha}^*$$

where

$$\alpha = \frac{1}{2} + \frac{1}{2}(\rho_0 - \rho_1).$$

Proof. We have

$$\begin{aligned}
\tilde{\eta}(x) &= \Pr(\tilde{Y} = 1 | X = x) \\
&= \Pr(\tilde{Y} = 1 | Y = 1, X = x) \Pr(Y = 1 | X = x) \\
&\quad + \Pr(\tilde{Y} = 1 | Y = 0, X = x) \Pr(Y = 0 | X = x) \\
&= (1 - \rho_1)\eta(x) + \rho_0(1 - \eta(x)) \\
&= (1 - \rho_0 - \rho_1)\eta(x) + \rho_0.
\end{aligned}$$

*Main article 10.1214/16-EJS1193.

It follows that

$$\eta(x) - \frac{1}{2} = \frac{\tilde{\eta}(x) - \alpha}{1 - \rho_0 - \rho_1}$$

where

$$\alpha = \frac{1}{2} + \frac{1}{2}(\rho_0 - \rho_1),$$

and therefore

$$\begin{aligned}(1-\rho_0-\rho_1)(R_P(f) - R_P^*) &= (1-\rho_0-\rho_1)\mathbb{E}_X[|\eta(X) - \tfrac{1}{2}|\mathbf{1}_{\{u(f(X)) \neq u(\eta(X) - \frac{1}{2})\}}] \\ &= \mathbb{E}_X[|\tilde{\eta}(X) - \alpha|\mathbf{1}_{\{u(f(X)) \neq u(\tilde{\eta}(X) - \alpha)\}}] \\ &= R_{\tilde{P},\alpha}(f) - R^*_{\tilde{P},\alpha}\end{aligned}$$

where we have used Eqns. (20) and (21) from [1] (Note that the left-hand side of Eqn. (21) should be $R_{P,\alpha}(f) - R^*_{P,\alpha}$). □

Remark 1. *$\pi_0 + \pi_1 < 1$ is equivalent to $\rho_0 + \rho_1 < 1$ provide $q := \Pr(Y = 1)$ and $\tilde{q} := \Pr(\tilde{Y} = 1)$ satisfy $0 < q, \tilde{q} < 1$. To see this, suppose $\rho_0 + \rho_1 < 1$. Bayes' rule gives*

$$\pi_0 = \frac{\rho_1 q}{\rho_1 q + (1-\rho_0)(1-q)}$$

and

$$\pi_1 = \frac{\rho_0(1-q)}{\rho_0(1-q) + (1-\rho_1)q},$$

and algebra leads to

$$1 - \pi_0 - \pi_1 = (1-\rho_0-\rho_1)\frac{q(1-q)}{[\rho_1 q + (1-\rho_0)(1-q)][\rho_0(1-q) + (1-\rho_1)q]}.$$

The ratio on the right is positive provided $0 < q < 1$, and thus $\pi_0 + \pi_1 < 1$. The reverse implication uses identical reasoning, and establishes that if $\pi_0 + \pi_1 < 1$, then

$$\rho_0 = \frac{\pi_1 \tilde{q}}{\pi_1 \tilde{q} + (1-\pi_0)(1-\tilde{q})}$$

and

$$\rho_1 = \frac{\pi_0(1-\tilde{q})}{\pi_0(1-\tilde{q}) + (1-\pi_1)\tilde{q}}$$

and $\rho_0 + \rho_1 < 1$. In light of this remark, it should be stipulated that $0 < q, \tilde{q} < 1$ throughout Section 7.

Remark 2. *The remainder of the arguments in Section 7 now carry forward with this new α. The only difference is that a different estimator for $\alpha = \frac{1}{2} + \frac{1}{2}(\rho_0 - \rho_1)$ is needed. Such an estimator is obtained from the formulas for ρ_0 and ρ_1 in terms of π_0, π_1, and $\tilde{q}$ in the previous remark. Since $\tilde{q}$ can be estimated easily from the noisy training data, and π_0 and π_1 can be estimated in the same way as described in the paper, we still get a consistent estimator for α, and this estimator will still satisfy Proposition 17 (the rate of convergence of $\widehat{\alpha}$ to α) under* **(C')**. *Thus, the proofs of the main consistency results still hold.*

Remark 3. *The alternate lemma described above actually allows us to drop the condition* **(A')** *in the consistency results, and return to the weaker condition* **(A)**.

Remark 4. *In light of Remark 1 above, the assumption in the co-training theorem of Section 8 can be replaced by the equivalent condition that the sum of the false positive and false negative rates of h is less than 1. In light of Remark 3, the last sentence of Section 8 should be modified so that assuming* **(A')** *is no longer required.*

References

[1] G. Blanchard, M. Flaska, G. Handy, S. Pozzi, and C. Scott. Classification with asymmetric label noise: Consistency and maximal denoising. *Electronic Journal of Statistics*, 10:2780–2824, 2016. MR3549019

[2] N. Natarajan, I. S. Dhillon, P. Ravikumar, and A. Tewari. Learning with noisy labels. In *Advances in Neural Information Processing Systems 26*, 2013.

[3] N. Natarajan, I. S. Dhillon, P. Ravikumar, and A. Tewari. Cost-sensitive learning with noisy labels. *Journal of Machine Learning Research*, 18 (155):1–33, 2018. URL `http://jmlr.org/papers/v18/15-226.html`.