Supplementary Material
Can Less be More? When Increasing-to-Balancing Label Noise Rates Considered Beneficial

Yang Liu
Computer Science and Engineering
University of California, Santa Cruz
Santa Cruz, CA 95064
yangliu@ucsc.edu

Jialu Wang
Computer Science and Engineering
University of California, Santa Cruz
Santa Cruz, CA 95064
faldict@ucsc.edu

This Appendix is organized as follows:

- Section A includes omitted proofs for theoretical conclusions in the main paper, as well as the extension to fairness constrained setting (A.9) and multi-class classification (A.10).
- Section B presents more experimental details and results.

A Omitted Proofs

A.1 Proof for Theorem 1

Proof Let $\ell^\diamond$ denote the noise-corrected loss with respect to true noise parameters e_+, e_-:

$$\ell^\diamond(h(x_n), \tilde{y}_n) := (1 - e_{-sgn(\tilde{y}_n)}) \cdot \ell(h(x_n), \tilde{y}_n) - e_{sgn(\tilde{y}_n)} \cdot \ell(h(x_n), -\tilde{y}_n) \tag{A1}$$

It was established in [25] the unbiasedness of $\ell^\diamond$:

Lemma 11 (Unbiasedness of $\ell^\diamond$, [25]). $\frac{1}{1-e_+-e_-} \cdot \mathbb{E}_{\tilde{Y}|Y=y}[\ell^\diamond(h(x), \tilde{Y})] = \ell(h(x), y)$.

A direct consequence of this lemma, via repeatedly applying to each (X, Y), is its unbiasedness on the population level:

$$\frac{1}{1-e_+-e_-} \cdot \mathbb{E}_{\tilde{D}|D}[\hat{R}_{\ell^\diamond,\tilde{D}}(h)] = \hat{R}_{\ell,D}(h), \quad \frac{1}{1-e_+-e_-} \cdot R_{\ell^\diamond,\tilde{\mathcal{D}}}(h) = R_{\ell,\mathcal{D}}(h)$$

The following fact holds by subtracting $\ell^\diamond$ from $\tilde{\ell}$:

$$\tilde{\ell}(h(x_n), \tilde{y}_n) = \ell^\diamond(h(x_n), \tilde{y}_n) + (e_{-\tilde{y}_n} - \tilde{e}_{-\tilde{y}_n}) \cdot \ell(h(x_n), \tilde{y}_n) + (\tilde{e}_{-\tilde{y}_n} - e_{\tilde{y}_n}) \cdot \ell(h(x_n), -\tilde{y}_n)$$

Using the triangle inequality of $|\cdot|$ we establish that

$$|\tilde{\ell}(h(x_n), \tilde{y}_n) - \ell^\diamond(h(x_n), \tilde{y}_n)| \leq \max\{|\tilde{e}_+ - e_+|, |\tilde{e}_- - e_-|\} \cdot \bar{\ell}. \tag{A2}$$

This further helps us bound the differences in the empirical risks:

$$|\hat{R}_{\tilde{\ell},\tilde{D}}(h) - \hat{R}_{\ell^\diamond,\tilde{D}}(h)| \leq \max\{|\tilde{e}_+ - e_+|, |\tilde{e}_- - e_-|\} \cdot \bar{\ell} \tag{A3}$$

Therefore

$$\begin{aligned}
\hat{R}_{\ell^\diamond,\tilde{D}}(h^*_{\tilde{\ell},\tilde{D}}) &\leq \hat{R}_{\tilde{\ell},\tilde{D}}(h^*_{\tilde{\ell},\tilde{D}}) + \max\{|\tilde{e}_+ - e_+|, |\tilde{e}_- - e_-|\} \cdot \bar{\ell} \\
&\leq \hat{R}_{\tilde{\ell},\tilde{D}}(h^*_{\ell,\mathcal{D}}) + \max\{|\tilde{e}_+ - e_+|, |\tilde{e}_- - e_-|\} \cdot \bar{\ell} && \text{(Opt. of } h^*_{\ell,\tilde{D}}\text{)} \\
&\leq \hat{R}_{\ell^\diamond,\tilde{D}}(h^*_{\ell,\mathcal{D}}) + 2\max\{|\tilde{e}_+ - e_+|, |\tilde{e}_- - e_-|\} \cdot \bar{\ell} && \text{(A4)}
\end{aligned}$$

Calling the results in [25], [Rademacher bound for max risk deviation, Proof of Lemma 2 therein], we know that for any $\delta > 0$, with probability at least $1-\delta$:

$$\sup_{h\in\mathcal{H}} \frac{1}{1-e_+-e_-}\left|R_{\ell^\diamond,\tilde{\mathcal{D}}}(h) - \hat{R}_{\ell^\diamond,\tilde{D}}(h)\right| \le \frac{2L}{1-e_+-e_-}\cdot\mathcal{R}(\mathcal{H}) + \sqrt{\frac{\log 1/\delta}{2N}} \tag{A5}$$

The above knowledge further helps us establish that

$$\begin{aligned}
& R_{\ell,\mathcal{D}}(h^*_{\tilde{\ell},\tilde{D}}) - R_{\ell,\mathcal{D}}(h^*_{\ell,\mathcal{D}}) \\
=\ & \frac{1}{1-e_+-e_-}(R_{\ell^\diamond,\tilde{\mathcal{D}}}(h^*_{\tilde{\ell},\tilde{D}}) - R_{\ell^\diamond,\tilde{\mathcal{D}}}(h^*_{\ell,\mathcal{D}})) && \text{(Unbiasedness of } \ell^\diamond \text{ on } \tilde{\mathcal{D}}) \\
=\ & \frac{1}{1-e_+-e_-}(R_{\ell^\diamond,\tilde{\mathcal{D}}}(h^*_{\tilde{\ell},\tilde{D}}) - \hat{R}_{\ell^\diamond,\tilde{D}}(h^*_{\tilde{\ell},\tilde{D}})) && \text{(Rademacher bound)} \\
& + \frac{1}{1-e_+-e_-}(\hat{R}_{\ell^\diamond,\tilde{D}}(h^*_{\tilde{\ell},\tilde{D}}) - \hat{R}_{\ell^\diamond,\tilde{D}}(h^*_{\ell,\mathcal{D}})) && \text{(Eqn. (A4))} \\
& + \frac{1}{1-e_+-e_-}(\hat{R}_{\ell^\diamond,\tilde{D}}(h^*_{\ell,\mathcal{D}}) - R_{\ell^\diamond,\tilde{\mathcal{D}}}(h^*_{\ell,\mathcal{D}})) && \text{(Rademacher bound)} \\
\le\ & \frac{4L}{1-e_+-e_-}\cdot\mathcal{R}(\mathcal{H}) + 2\sqrt{\frac{\log 1/\delta}{2N}} + 2\frac{\max\{|\tilde{e}_+-e_+|, |\tilde{e}_--e_-|\}}{1-e_+-e_-}\cdot\bar{\ell} \\
\le\ & \frac{4L}{1-e_+-e_-}\cdot\mathcal{R}(\mathcal{H}) + 2\sqrt{\frac{\log 1/\delta}{2N}} + 2\frac{\texttt{err}_{\texttt{M}}}{1-e_+-e_-}\cdot\bar{\ell}\,.
\end{aligned}$$

We complete the proof. ■

A.2 Proof for Lemma 2

Proof Expanding $\mathbb{P}(h(X)\neq\hat{Y})$ using the law of total probability we have

$$\begin{aligned}
\mathbb{P}(h(X)\neq\hat{Y}) =\ & \mathbb{P}(h(X)\neq\hat{Y}, \hat{Y}\neq Y) + \mathbb{P}(h(X)\neq\hat{Y}, \hat{Y}= Y) \\
=\ & \mathbb{P}(h(X)\neq\hat{Y}\mid\hat{Y}\neq Y)\cdot\mathbb{P}(\hat{Y}\neq Y) + \mathbb{P}(h(X)\neq\hat{Y}\mid\hat{Y}= Y)\cdot\mathbb{P}(\hat{Y}= Y).
\end{aligned}$$

In binary classification, $h(X)\neq\tilde{Y}, \tilde{Y}\neq Y$ implies $h(X)=Y$, such that

$$\mathbb{P}(h(X)\neq\tilde{Y}\mid\tilde{Y}\neq Y) = \mathbb{P}(h(X)= Y\mid\tilde{Y}\neq Y).$$

Due to the independence of $\tilde{Y}$ and X given Y,

$$\mathbb{P}(h(X)= Y\mid\tilde{Y}\neq Y) = \frac{\mathbb{P}(h(X)= Y, \tilde{Y}\neq Y)}{\mathbb{P}(\tilde{Y}\neq Y)} = \frac{\mathbb{P}(h(X)= Y)\mathbb{P}(\tilde{Y}\neq Y)}{\mathbb{P}(\tilde{Y}\neq Y)} = \mathbb{P}(h(X)= Y)$$

Similarly, we have

$$\mathbb{P}(h(X)\neq\hat{Y}\mid\hat{Y}= Y) = \mathbb{P}(h(X)\neq\tilde{Y}).$$

Combining all above we finished the proof when $e<0.5$ by having:

$$\begin{aligned}
\mathbb{P}(h(X)\neq\hat{Y}) =\ & \mathbb{P}(h(X)= Y)\cdot e + \mathbb{P}(h(X)\neq Y)\cdot(1-e) \\
=\ & (1-2e)\cdot\mathbb{P}(h(X)\neq Y) + e
\end{aligned}$$

■

A.3 Proof for Theorem 3

Proof Again let $\ell^\diamond$ denote the noise-corrected loss with respect to true noise parameters e_+, e_-:

$$\ell^\diamond(h(x_n), \tilde{y}_n) := (1-e_{-sgn(\tilde{y}_n)})\cdot\ell(h(x_n), \tilde{y}_n) - e_{sgn(\tilde{y}_n)}\cdot\ell(h(x_n), -\tilde{y}_n) \tag{A6}$$

First notice the following when ℓ is a symmetric loss:

$$\begin{aligned}
& R_{\ell,\mathcal{D}}(h^*_{\ell,\hat{D}}) \\
=\ & \frac{1}{1-2e}\cdot R_{\ell^\diamond,\hat{\mathcal{D}}}(h^*_{\ell,\hat{D}}) && \text{(Unbiasedness of } \ell^\diamond \text{ on } \hat{\mathcal{D}} \text{ using symmetric } e) \\
=\ & \frac{1-e}{1-2e}\cdot R_{\ell,\hat{\mathcal{D}}}(h^*_{\ell,\hat{D}}) - \frac{e}{1-2e}\cdot R_{\ell,\hat{\mathcal{D}}}(-h^*_{\ell,\hat{D}}) && \text{(A7)}
\end{aligned}$$

The last equality uses the definition of $\ell^\diamond$, and is due to ℓ being symmetric. Then we show that

$$\begin{aligned}
&\frac{1}{1-2e}\cdot\left(R_{\ell,\hat{\mathcal{D}}}(h^*_{\ell,\hat{D}})-R_{\ell,\hat{\mathcal{D}}}(h^*_{\ell,\mathcal{D}})\right)\\
=\;&\frac{1}{1-2e}\left(R_{\ell,\hat{\mathcal{D}}}(h^*_{\ell,\hat{D}})-\hat{R}_{\ell,\hat{D}}(h^*_{\ell,\hat{D}})\right) && \text{(Rademacher bound)}\\
&+\frac{1}{1-2e}\left(\hat{R}_{\ell,\hat{D}}(h^*_{\ell,\hat{D}})-\hat{R}_{\ell,\hat{D}}(h^*_{\ell,\mathcal{D}})\right) && (\leq 0 \text{ Optimality of } h^*_{\ell,\hat{D}} \text{ on } \ell,\hat{D})\\
&+\frac{1}{1-2e}\left(\hat{R}_{\ell,\hat{D}}(h^*_{\ell,\mathcal{D}})-R_{\ell,\hat{\mathcal{D}}}(h^*_{\ell,\mathcal{D}})\right) && \text{(Rademacher bound)}\\
\leq\;&\frac{4L}{1-2e}\mathcal{R}(\mathcal{H})+2\sqrt{\frac{\log 1/\delta}{2N}}
\end{aligned}$$

The inequality is due to the Rademacher bound we invoked as in Eqn. (A5) as well as the optimality of $h^*_{\ell,\hat{D}}$ on $\ell,\hat{D}$. That is we have proved with probability at least $1-\delta$ that

$$\frac{1}{1-2e}\cdot R_{\ell,\hat{\mathcal{D}}}(h^*_{\ell,\hat{D}})\leq\frac{1}{1-2e}\cdot R_{\ell,\hat{\mathcal{D}}}(h^*_{\ell,\mathcal{D}})+\frac{4L}{1-2e}\mathcal{R}(\mathcal{H})+2\sqrt{\frac{\log 1/\delta}{2N}} \tag{A8}$$

Repeating the same analysis and using the assumed condition that $\hat{R}_{\ell,\hat{D}}(-h^*_{\ell,\hat{D}})-\hat{R}_{\ell,\hat{D}}(-h^*_{\ell,\mathcal{D}})\geq 0$ we have

$$\frac{1}{1-2e}\cdot R_{\ell,\hat{\mathcal{D}}}(-h^*_{\ell,\hat{D}})\geq\frac{1}{1-2e}\cdot R_{\ell,\hat{\mathcal{D}}}(-h^*_{\ell,\mathcal{D}})-\frac{4L}{1-2e}\mathcal{R}(\mathcal{H})-2\sqrt{\frac{\log 1/\delta}{2N}} \tag{A9}$$

Combining above (Eqn. (A8) and (A9)), we have with probability at least $1-\delta$ (that both of the above bounds will happen simultaneously)

$$\begin{aligned}
R_{\ell,\mathcal{D}}(h^*_{\ell,\hat{D}})=\;&\frac{1-e}{1-2e}\cdot R_{\ell,\hat{\mathcal{D}}}(h^*_{\ell,\hat{D}})-\frac{e}{1-2e}\cdot R_{\ell,\hat{\mathcal{D}}}(-h^*_{\ell,\hat{D}})\\
\leq\;&\frac{1-e}{1-2e}\cdot R_{\ell,\hat{\mathcal{D}}}(h^*_{\ell,\mathcal{D}})-\frac{e}{1-2e}\cdot R_{\ell,\hat{\mathcal{D}}}(-h^*_{\ell,\mathcal{D}})+\frac{4L}{1-2e}\mathcal{R}(\mathcal{H})+2\sqrt{\frac{\log 1/\delta}{2N}}\\
=\;&R_{\ell,\mathcal{D}}(h^*_{\ell,\mathcal{D}})+\frac{4L}{1-2e}\mathcal{R}(\mathcal{H})+2\sqrt{\frac{\log 1/\delta}{2N}}.
\end{aligned}$$

The inequality uses Eqn. (A8) and (A9). Again the last equality is reusing Eqn. (A7). This completes the proof. ■

A.4 Proof for Lemma 4

Proof

$$\mathbb{P}(h(X)=+1|\tilde{Y}=+1,Z=a)=\frac{\mathbb{P}(h(X)=+1,\tilde{Y}=+1|Z=a)}{\mathbb{P}(\tilde{Y}=+1|Z=a)} \tag{A10}$$

Again we do the trick of sampling $\mathbb{P}(\tilde{Y}=+1|Z=a)$ to be 0.5, which allows us to focus on the numerator.

$$\begin{aligned}
&\mathbb{P}(h(X)=+1,\tilde{Y}=+1|Z=a)\\
=\;&\mathbb{P}(h(X)=+1,\tilde{Y}=+1,Y=+1|Z=a)\\
&+\mathbb{P}(h(X)=+1,\tilde{Y}=+1,Y=-1|Z=a)\\
=\;&\mathbb{P}(h(X)=+1,\tilde{Y}=+1|Y=+1,Z=a)\cdot\mathbb{P}(Y=+1|Z=a)\\
&+\mathbb{P}(h(X)=+1,\tilde{Y}=+1|Y=-1,Z=a)\cdot\mathbb{P}(Y=-1|Z=a)\\
=\;&\mathbb{P}(h(X)=+1|Y=+1,Z=a)\cdot(1-e_a)\cdot\mathbb{P}(Y=+1|Z=a)\\
&+\mathbb{P}(h(X)=+1|Y=-1,Z=a)\cdot e_a\cdot\mathbb{P}(Y=-1|Z=a)\\
&\qquad\qquad\text{(Independence of } X \text{ and } \tilde{Y} \text{ given } Y)
\end{aligned}$$

That is

$$0.5 \cdot \widetilde{\text{TPR}}_a(h) = \text{TPR}_a(h) \cdot (1 - e_a) \cdot \mathbb{P}(Y = +1|Z = a) + \text{FPR}_a(h) \cdot e_a \cdot \mathbb{P}(Y = -1|Z = a) \tag{A11}$$

Similarly for FPR we have

$$\mathbb{P}(h(X) = +1|\tilde{Y} = -1, Z = a) = \frac{\mathbb{P}(h(X) = +1, \tilde{Y} = -1|Z = a)}{\mathbb{P}(\tilde{Y} = -1|Z = a)} \tag{A12}$$

Following similar steps as above, the numerator further derives as

$$\begin{aligned} & \mathbb{P}(h(X) = +1, \tilde{Y} = +1|Z = a) \\ = \ & \mathbb{P}(h(X) = +1|Y = -1, Z = a) \cdot (1 - e_a) \cdot \mathbb{P}(Y = +1|Z = a) \\ & + \mathbb{P}(h(X) = +1|Y = +1, Z = a) \cdot e_a \cdot \mathbb{P}(Y = -1|Z = a) \end{aligned}$$

That is

$$0.5 \cdot \widetilde{\text{FPR}}_a(h) = \text{FPR}_a(h) \cdot (1 - e_a) \cdot \mathbb{P}(Y = +1|Z = a) + \text{TPR}_a(h) \cdot e_a \cdot \mathbb{P}(Y = -1|Z = a) \tag{A13}$$

When $\mathbb{P}(\tilde{Y} = +1|Z = a) = \mathbb{P}(\tilde{Y} = +1|Z = b) = 0.5$, we will also have

$$0.5 = \mathbb{P}(\tilde{Y} = +1|Z = a) = \mathbb{P}(Y = +1|Z = a)(1 - e_a) + \mathbb{P}(Y = -1|Z = a)e_a \tag{A14}$$

which returns us that $\mathbb{P}(Y = +1|Z = a) = \frac{0.5 - e_a}{1 - 2e_a} := p = 0.5$. Using this knowledge and solving the linear equations defined by Eqn. (A11) and (A13) we have

$$\text{TPR}_a(h) = \frac{C_{a,1} \cdot \widetilde{\text{TPR}}_a(h) - C_{a,2} \cdot \widetilde{\text{FPR}}_a(h)}{e_a - 0.5} \tag{A15}$$

$$\text{FPR}_a(h) = \frac{C_{a,1} \cdot \widetilde{\text{FPR}}_a(h) - C_{a,2} \cdot \widetilde{\text{TPR}}_a(h)}{e_a - 0.5} \tag{A16}$$

■

A.5 Proof for Theorem 5

Proof Combining Eqn. (5) and (6) we have

$$\begin{aligned} & |\text{TPR}_z(h) - \text{TPR}_z^c(h)| \\ = \ & \left| \frac{0.5 \cdot e_z \cdot \widetilde{\text{TPR}}_z(h) - 0.5(1 - e_z) \cdot \widetilde{\text{FPR}}_z(h)}{e_z - 0.5} - \frac{0.5 \cdot \tilde{e}_z \cdot \widetilde{\text{TPR}}_z(h) - 0.5(1 - \tilde{e}_z) \cdot \widetilde{\text{FPR}}_z(h)}{\tilde{e}_z - 0.5} \right| \\ = \ & \frac{|\tilde{e}_z - e_z| \cdot \widetilde{\text{TPR}}_z(h)}{(2e_z - 1)(2\tilde{e}_z - 1)} \\ = \ & \frac{\texttt{err}_z \cdot \widetilde{\text{TPR}}_z(h)}{(2e_z - 1)(2\tilde{e}_z - 1)} \ . \end{aligned} \tag{A17}$$

Recall $\texttt{err}_z = |\tilde{e}_z - e_z|$. The second equality is algebraic - we simply unify the denominator of both quantities and rearrange terms. Then equalizing TPR that $\text{TPR}_a^c(h) = \text{TPR}_b^c(h)$ returns us

$$\begin{aligned} & |\text{TPR}_a(h) - \text{TPR}_b(h)| \\ = & |\text{TPR}_a(h) - \text{TPR}_a^c(h) + \text{TPR}_b^c(h) - \text{TPR}_b(h)| \\ \geq & \ ||\text{TPR}_a(h) - \text{TPR}_a^c(h)| - |\text{TPR}_b^c(h) - \text{TPR}_b(h)|| \\ = & \left| \frac{\texttt{err}_a \cdot \widetilde{\text{TPR}}_a(h)}{(2e_a - 1)(2\tilde{e}_a - 1)} - \frac{\texttt{err}_b \cdot \widetilde{\text{TPR}}_b(h)}{(2e_b - 1)(2\tilde{e}_b - 1)} \right|, \end{aligned}$$

where the last equality is an application of Eqn. (A17). Then

$$\begin{aligned}
&\left|\frac{\texttt{err}_a \cdot \widetilde{\text{TPR}}_a(h)}{(2e_a-1)(2\tilde{e}_a-1)} - \frac{\texttt{err}_b \cdot \widetilde{\text{TPR}}_b(h)}{(2e_b-1)(2\tilde{e}_b-1)}\right| \\
=&\texttt{err}_a \cdot \left|\frac{\widetilde{\text{TPR}}_a(h)}{(2e_a-1)(2\tilde{e}_a-1)} - \frac{\texttt{err}_b}{\texttt{err}_a}\frac{\widetilde{\text{TPR}}_b(h)}{(2e_b-1)(2\tilde{e}_b-1)}\right| \\
\geq&\texttt{err}_{\texttt{M}} \cdot \left|\frac{\widetilde{\text{TPR}}_a(h)}{(2e_a-1)(2\tilde{e}_a-1)} - \frac{\texttt{err}_b}{\texttt{err}_a}\frac{\widetilde{\text{TPR}}_b(h)}{(2e_b-1)(2\tilde{e}_b-1)}\right|
\end{aligned}$$

Similarly

$$\begin{aligned}
&|\text{FPR}_z(h) - \text{FPR}^c_z(h)| \\
=& \left|\frac{0.5\cdot e_z \cdot \widetilde{\text{FPR}}_z(h) - 0.5(1-e_z)\cdot\widetilde{\text{TPR}}_z(h)}{e_z - 0.5} - \frac{0.5\cdot \tilde{e}_z \cdot \widetilde{\text{FPR}}_z(h) - 0.5(1-\tilde{e}_z)\cdot\widetilde{\text{TPR}}_z(h)}{\tilde{e}_z - 0.5}\right| \\
=& \frac{|\tilde{e}_z - e_z|\cdot\widetilde{\text{FPR}}_z(h)}{(2e_z-1)(2\tilde{e}_z-1)}. \\
=& \frac{\texttt{err}_z\cdot\widetilde{\text{FPR}}_z(h)}{(2e_z-1)(2\tilde{e}_z-1)}.
\end{aligned}$$

Then equalizing FPR that $\text{FPR}^c_a(h) = \text{FPR}^c_b(h)$ we have

$$\begin{aligned}
&|\text{FPR}_a(h) - \text{FPR}_b(h)| \\
=&|\text{FPR}_a(h) - \text{FPR}^c_a(h) + \text{FPR}^c_b(h) - \text{FPR}_b(h)| \\
\geq&\,||\text{FPR}_a(h) - \text{FPR}^c_a(h)| - |\text{FPR}^c_b(h) - \text{FPR}_b(h)|| \\
\geq&\left|\frac{\texttt{err}_a\cdot\widetilde{\text{FPR}}_a(h)}{(2e_a-1)(2\tilde{e}_a-1)} - \frac{\texttt{err}_b\cdot\widetilde{\text{FPR}}_b(h)}{(2e_b-1)(2\tilde{e}_b-1)}\right| \\
\geq&\texttt{err}_{\texttt{M}}\cdot\left|\frac{\widetilde{\text{FPR}}_a(h)}{(2e_a-1)(2\tilde{e}_a-1)} - \frac{\texttt{err}_b}{\texttt{err}_a}\frac{\widetilde{\text{FPR}}_b(h)}{(2e_b-1)(2\tilde{e}_b-1)}\right|.
\end{aligned}$$

■

A.6 Proof for Theorem 6

Proof Easy to show that when $e_a = e_b$, $C_{a,1} = C_{b,1}$ and $C_{a,2} = C_{b,2}$. Therefore, from Eqn. (5) we know equalizing

$$\widetilde{\text{TPR}}_a(h) = \widetilde{\text{TPR}}_b(h), \;\; \widetilde{\text{FPR}}_a(h) = \widetilde{\text{FPR}}_b(h) \tag{A18}$$

will also return us

$$\text{TPR}_a(h) = \text{TPR}_b(h), \;\; \text{FPR}_a(h) = \text{FPR}_b(h) \tag{A19}$$

■

A.7 Proof for Theorem 9

Proof We start with deriving $\text{PA}_{\mathcal{D}^\diamond}$.

$$\text{PA}_{\mathcal{D}^\diamond} = \mathbb{P}(\tilde{Y}_2 = \tilde{Y}_3 = +1 | \tilde{Y}_1 = +1) = \frac{\mathbb{P}(\tilde{Y}_1 = \tilde{Y}_2 = \tilde{Y}_3 = +1)}{\mathbb{P}(\tilde{Y}_1 = +1)}$$

Due to the sampling step, we have $\mathbb{P}(\tilde{Y}_1 = +1) = 0.5$ - this allows us to focus on the denominator:

$$\begin{aligned}
\mathbb{P}(\tilde{Y}_1 = \tilde{Y}_2 = \tilde{Y}_3 = +1) &\overset{(1)}{=} \mathbb{P}(Y=+1)\prod_{i=1}^3 \mathbb{P}(\tilde{Y}_i = +1 | Y = +1) + \mathbb{P}(Y=-1)\prod_{i=1}^3 \mathbb{P}(\tilde{Y}_i = +1 | Y = -1) \\
&\overset{(2)}{=} \mathbb{P}(Y=+1)\cdot(1-e_+)^3 + \mathbb{P}(Y=-1)\cdot e_-^3
\end{aligned}$$

where in above, (1) uses the 2-NN clusterability of D, and (2) uses the conditional independence between the noisy labels. Similarly for $\text{NA}_{\mathcal{D}^\diamond}$ we have:

$$\mathbb{P}(\tilde{Y}_2 = \tilde{Y}_3 = -1 | \tilde{Y}_1 = -1) = \frac{\mathbb{P}(\tilde{Y}_1 = \tilde{Y}_2 = \tilde{Y}_3 = -1)}{\mathbb{P}(\tilde{Y}_1 = -1)}$$

Again we have that $\mathbb{P}(\tilde{Y}_1 = -1) = 0.5$, and the numerator derives as

$$\begin{aligned}\mathbb{P}(\tilde{Y}_1 = \tilde{Y}_2 = \tilde{Y}_3 = -1) &= \mathbb{P}(Y = +1) \prod_{i=1}^{3} \mathbb{P}(\tilde{Y}_i = -1 | Y = +1) + \mathbb{P}(Y = -1) \prod_{i=1}^{3} \mathbb{P}(\tilde{Y}_i = -1 | Y = -1) \\ &= \mathbb{P}(Y = +1) \cdot e_+^3 + \mathbb{P}(Y = -1) \cdot (1 - e_-)^3\end{aligned}$$

Taking the difference (and normalize by 0.5) we have

$$\begin{aligned}& 0.5 \cdot (\text{PA}_{\mathcal{D}^\diamond} - \text{NA}_{\mathcal{D}^\diamond}) \\ =\ & \mathbb{P}(\tilde{Y}_2 = \tilde{Y}_3 = +1 | \tilde{Y}_1 = +1) - \mathbb{P}(\tilde{Y}_2 = \tilde{Y}_3 = -1 | \tilde{Y}_1 = -1) \\ =\ & \mathbb{P}(Y = +1) \left((1 - e_+)^3 - e_+^3 \right) + \mathbb{P}(Y = -1) \left(e_-^3 - (1 - e_-)^3 \right) \end{aligned} \tag{A20}$$

Notice two facts: first we can derive that

$$(1 - e_+)^3 - e_+^3 = (1 - 2e_+)(e_+^2 - e_+ + 1), \ \ e_-^3 - (1 - e_-)^3 = -(1 - 2e_-)(e_-^2 - e_- + 1)$$

Second, we will use the following fact:

$$0.5 = \mathbb{P}(\tilde{Y} = +1) = \mathbb{P}(Y = +1)(1 - e_+) + \mathbb{P}(Y = -1) e_- \tag{A21}$$

from which we solve that $\mathbb{P}(Y = +1) = \frac{0.5 - e_-}{1 - e_+ - e_-}$. Symmetrically, $\mathbb{P}(Y = -1) = \frac{0.5 - e_+}{1 - e_+ - e_-}$.

Return the above two facts back into Eqn. (A20), we have

$$\begin{aligned}& \mathbb{P}(Y = +1)((1 - e_+)^3 - e_+^3) + \mathbb{P}(Y = -1)(e_-^3 - (1 - e_-)^3) \\ =\ & 2 \cdot \frac{(0.5 - e_+)(0.5 - e_-)}{1 - e_+ - e_-} \left((e_+^2 - e_+ + 1) - (e_-^2 - e_- + 1) \right) \\ =\ & 2 \cdot (0.5 - e_+) \cdot (0.5 - e_-) \cdot (e_- - e_+)\end{aligned}$$

completing the proof when $e_+, e_- < 0.5$. ■

A.8 Proof for Proposition 10

Proof Expanding $\mathbb{P}(\hat{Y} = -1 | Y = +1)$ using the law of total probability we have

$$\begin{aligned}\hat{e}_+ &= \mathbb{P}(\hat{Y} = -1 | Y = +1) \\ &= \mathbb{P}(\hat{Y} = -1, \tilde{Y} = +1 | Y = +1) + \mathbb{P}(\hat{Y} = -1, \tilde{Y} = -1 | Y = +1) \\ &= \mathbb{P}(\hat{Y} = -1 | \tilde{Y} = +1, Y = +1) \cdot \mathbb{P}(\tilde{Y} = +1 | Y = +1) \\ &\quad + \mathbb{P}(\hat{Y} = -1 | \tilde{Y} = -1, Y = +1) \cdot \mathbb{P}(\tilde{Y} = -1 | Y = +1) \\ &= \epsilon \cdot (1 - e_+) + 1 \cdot e_+ \qquad \text{(Independence between } \hat{Y} \text{ and } Y \text{ given } \tilde{Y}) \\ &= (1 - e_+) \cdot \epsilon + e_+\end{aligned}$$

Similarly,

$$\begin{aligned}\hat{e}_- &= \mathbb{P}(\hat{Y} = +1 | Y = -1) \\ &= \mathbb{P}(\hat{Y} = +1, \tilde{Y} = +1 | Y = -1) + \mathbb{P}(\hat{Y} = +1, \tilde{Y} = -1 | Y = -1) \\ &= \mathbb{P}(\hat{Y} = +1 | \tilde{Y} = +1, Y = -1) \cdot \mathbb{P}(\tilde{Y} = +1 | Y = -1) \\ &\quad + \mathbb{P}(\hat{Y} = +1 | \tilde{Y} = -1, Y = -1) \cdot \mathbb{P}(\tilde{Y} = -1 | Y = -1) \\ &= (1 - \epsilon) \cdot e_-.\end{aligned}$$

The last equality is again due to the independence between $\hat{Y}$ and Y given $\tilde{Y}$, as well as the fact that we do not flip the $\tilde{Y} = -1$ labels so $\mathbb{P}(\hat{Y} = +1 | \tilde{Y} = -1, Y = -1) = 0$. Taking the difference we finish the proof. ■

A.9 Balancing noise for fairness constrained case

Define

$$\text{PA}_{\mathcal{D}^\diamond,a} = \mathbb{P}_{Z=a}(\tilde{Y}_2 = \tilde{Y}_3 = +1 | \tilde{Y}_1 = +1) \tag{A22}$$

$$\text{PA}_{\mathcal{D}^\diamond,b} = \mathbb{P}_{Z=b}(\tilde{Y}_2 = \tilde{Y}_3 = +1 | \tilde{Y}_1 = +1) \tag{A23}$$

We now claim that $sgn(\text{PA}_{\mathcal{D}^\diamond,a} - \text{PA}_{\mathcal{D}^\diamond,b}) = -sgn(e_a - e_b)$. We start with deriving $\text{PA}_{\mathcal{D}^\diamond,a}$.

$$\text{PA}_{\mathcal{D}^\diamond,a} = \mathbb{P}_{Z=a}(\tilde{Y}_2 = \tilde{Y}_3 = +1 | \tilde{Y}_1 = +1) = \frac{\mathbb{P}_{Z=a}(\tilde{Y}_1 = \tilde{Y}_2 = \tilde{Y}_3 = +1)}{\mathbb{P}_{Z=a}(\tilde{Y}_1 = +1)}$$

Due to the sampling step, we have $\mathbb{P}_{Z=a}(\tilde{Y}_1 = +1) = 0.5$ - this allows us to focus on the denominator:

$$\begin{aligned}
\mathbb{P}_{Z=a}(\tilde{Y}_1 = \tilde{Y}_2 = \tilde{Y}_3 = +1) &\overset{(1)}{=} \mathbb{P}_{Z=a}(Y = +1) \prod_{i=1}^{3} \mathbb{P}_{Z=a}(\tilde{Y}_i = +1 | Y = +1) \\
&\quad + \mathbb{P}_{Z=a}(Y = -1) \prod_{i=1}^{3} \mathbb{P}_{Z=a}(\tilde{Y}_i = +1 | Y = -1) \\
&\overset{(2)}{=} \mathbb{P}_{Z=a}(Y = +1) \cdot (1 - e_a)^3 + \mathbb{P}_{Z=a}(Y = -1) \cdot e_a^3
\end{aligned}$$

where in above, (1) uses the 2-NN clusterability of D, and (2) uses the conditional independence between the noisy labels. Similarly for $\text{PA}_{\mathcal{D}^\diamond,b}$ we have:

$$\text{PA}_{\mathcal{D}^\diamond,b} = \frac{\mathbb{P}_{Z=b}(Y = +1) \cdot (1 - e_b)^3 + \mathbb{P}_{Z=b}(Y = -1) \cdot e_b^3}{0.5} \tag{A24}$$

Firstly, we will use the following fact for $z \in \{a, b\}$:

$$\begin{aligned}
0.5 =& \mathbb{P}_{Z=z}(\tilde{Y} = +1) \\
=& \mathbb{P}_{Z=z}(\tilde{Y} = +1 | Y = +1) \cdot \mathbb{P}_{Z=z}(Y = +1) + \mathbb{P}_{Z=z}(\tilde{Y} = +1 | Y = -1) \cdot \mathbb{P}_{Z=z}(Y = -1) \\
=& \mathbb{P}_{Z=z}(Y = +1) \cdot (1 - e_z) + \mathbb{P}_{Z=z}(Y = -1) \cdot e_z
\end{aligned}$$

from which we solve that $\mathbb{P}_{Z=z}(Y = +1) = \frac{0.5 - e_z}{1 - 2e_z} = 0.5$. Therefore

$$\begin{aligned}
\text{PA}_{\mathcal{D}^\diamond,a} - \text{PA}_{\mathcal{D}^\diamond,b} &= (1 - e_a)^3 - (1 - e_b)^3 + e_a^3 - e_b^3 \\
&= (e_b - e_a) \left((1 - e_a)^2 + (1 - e_b)^2 + (1 - e_a)(1 - e_b) - e_a^2 - e_b^2 - e_a e_b \right) \\
&= (e_b - e_a) \left(1 - 2e_a + 1 - 2e_b + 1 - e_a - e_b \right)
\end{aligned} \tag{A25}$$

Note that $1 - 2e_a + 1 - 2e_b + 1 - e_a - e_b > 0$ when $e_a, e_b < 0.5$. This implies that we can use the 2-NN positive agreements $\text{PA}_{\mathcal{D}^\diamond,a} - \text{PA}_{\mathcal{D}^\diamond,b}$ across groups to compare e_a with e_b.

A.10 Extension to multi-class

As explained at the beginning, our algorithm can largely extend to the multi-class/group setting. The primary requirement of the extension is to extend the definition of $\text{PA}_{\mathcal{D}^\diamond}, \text{NA}_{\mathcal{D}^\diamond}$ to each label class/group. Consider a multi-class classification problem with K label classes, and the noise rates follow a uniform diagonal model:

$$\mathbb{P}(\tilde{Y} = k | Y = k) = 1 - e_k, \;\; \mathbb{P}(\tilde{Y} = k' | Y = k) = \frac{e_k}{K - 1}, \; \forall k' \neq k. \tag{A26}$$

Define $\text{KA}_{\mathcal{D}^\diamond,k} := \mathbb{P}(\tilde{Y}_2 = \tilde{Y}_3 = k | \tilde{Y}_1 = k), \; k = 1, 2, ..., K$. Similarly we can show that for any pair of k_1, k_2: $sgn(\text{KA}_{\mathcal{D}^\diamond,k_1} - \text{KA}_{\mathcal{D}^\diamond,k_2}) = -sgn(e_{k_1} - e_{k_2})$, wherein above e_{k_1}, e_{k_2} are the error rates of label class k_1, k_2. With the above, we can compute $\text{KA}_{\mathcal{D}^\diamond,k}$, rank them, and start inserting noise to the classes that are determined to have a lower error rate to match the highest one.

A.11 Pseudocodes

```
import numpy as np
from sklearn.neighbors import NearestNeighbors
def estimate_PA(X, y):
    nbrs = NearestNei.(n_neighbors=3, algorithm='ball_tree').fit(X)
    _, indices = nbrs.kneighbors(X)
    return np.mean(np.array([np.all(y[i] == y[indices[i]]) for i in np
        .where(y > 0)[0]]))
```

Figure A1: **Numpy-like pseudocode for an implementation of estimating PA.** Our implementation utilizes scikit-learn's Nearest Neighbors module. The code for esimating NA is similar.

B Additional Experiment Details and Results

We provide more details on the experimental setup as well as further results.

B.1 Datasets

We evaluate our methods on five datasets:

- Adult, the UCI Adult Income dataset [9]. The task is to predict whether an individual's income exceeds 50K. The dataset consists of 48,842 examples and 28 features. We select female and male as two protected groups in constrained learning. We resample the dataset to ensure that both the classes and groups are balanced.
- Compas, the COMPAS recidivism dataset for crime statistics with 7,168 instances and 10 features [2]. We select race as the protected attribute in constrained learning.
- Fairface, the face attribute dataset containing 108,501 images with balanced race and gender groups [15]. We use a pre-trained vision transformer (ViT/B-32) model [8] to extract image representations, and project them into 50-dimensional feature vectors. For both unconstrained and constrained learning, we take gender attribute as labels for binary classification. For constrained learning, we categorize race into White and Non-White groups.
- MNIST [18], consisting of 50,000 training images and 10,000 test images in 10 classes. We train a MLP model from scratch on the MNIST dataset.
- CIFAR-10 [16], consisting of 50,000 training images and 10,000 test images in 10 classes. We evaluate unconstrained multi-class classification on CIFAR-10 dataset. Similar to Fairface, we use a pre-trained vision transformer to extract 512-dimensional feature vectors.

For Adult, Compas, and German datasets, we perform random train/test splits in a ratio of 80 to 20. For Fairface, MNIST, and CIFAR-10, we follow their original splits.

B.2 Computing infrastructure

For all the experiments, we use a GPU cluster with 4 2080 Ti GPUs for training and evaluation.

B.3 Noise transition matrix for CIFAR-10

We adopt the following procedure to generate the noise transition matrix:

1. Manually set the diagonal elements at least 0.4. We ensure that the difference between the maximal elements and 0.4 is equal to the noise gap.
2. Permute the diagonal elements to increase the randomness.
3. Fill out the non-diagonal elements randomly and ensure the sum of each column is 1

We show one sample noise transition matrix generated by our procedure with noise gap 0.2 as follows:

$$\begin{bmatrix}
\mathbf{0.4} & 0.087 & 0.013 & 0.032 & 0.032 & 0.068 & 0.050 & 0.178 & 0.001 & 0.118 \\
0.043 & \mathbf{0.4} & 0.002 & 0.016 & 0.049 & 0.113 & 0.060 & 0.024 & 0.224 & 0.017 \\
0.181 & 0.111 & \mathbf{0.4} & 0.147 & 0.033 & 0.005 & 0.026 & 0.040 & 0.110 & 0.076 \\
0.051 & 0.001 & 0.060 & \mathbf{0.6} & 0.032 & 0.047 & 0.149 & 0.145 & 0.022 & 0.059 \\
0.001 & 0.167 & 0.119 & 0.032 & \mathbf{0.6} & 0.092 & 0.051 & 0.018 & 0.037 & 0.129 \\
0.097 & 0.007 & 0.001 & 0.059 & 0.016 & \mathbf{0.4} & 0.019 & 0.014 & 0.084 & 0.001 \\
0.018 & 0.023 & 0.277 & 0.041 & 0.034 & 0.014 & \mathbf{0.4} & 0.028 & 0.041 & 0.062 \\
0.149 & 0.096 & 0.081 & 0.019 & 0.041 & 0.015 & 0.143 & \mathbf{0.4} & 0.061 & 0.110 \\
0.031 & 0.066 & 0.022 & 0.007 & 0.133 & 0.080 & 0.049 & 0.113 & \mathbf{0.4} & 0.025 \\
0.029 & 0.040 & 0.023 & 0.043 & 0.027 & 0.162 & 0.048 & 0.036 & 0.018 & \mathbf{0.4}
\end{bmatrix}$$

B.4 Additional results

Table B1: **Binary classification accuracy of compared methods on 3 datasets across different levels of noise rates.** Mis. SL: surrogate loss [25] with misspecified parameters. Est. SL: surrogate loss [25] with estimated parameters. CE: vanilla cross entropy. Peer: peer loss function [21]. All methods are trained with one-layer perceptron with the same hyper-parameters. For each noise setting, we average across 5 runs and report the mean and standard deviation. We find that the increasing-to-balancing can boost the vanilla cross entropy on all the noise settings.

			BASELINES (LESS NOISE)				NOISE+ (MORE NOISE)	
Dataset	e_-	e_+	Mis. SL	Est. SL	CE	Peer	CE	Peer
Adult	0.0	0.1	72.79 ± 0.34	72.64 ± 0.38	72.63 ± 0.29	72.77 ± 0.32	73.62 ± 0.37	73.86 ± 0.41
$n = 48,842$	0.0	0.2	72.27 ± 0.39	72.13 ± 0.37	71.26 ± 0.38	71.95 ± 0.34	72.73 ± 0.71	73.52 ± 0.58
$d = 28$	0.0	0.3	67.93 ± 0.52	71.58 ± 0.28	66.86 ± 0.47	71.33 ± 0.30	73.30 ± 0.27	73.74 ± 0.15
	0.0	0.4	55.54 ± 0.28	70.29 ± 0.28	63.97 ± 1.07	69.38 ± 0.41	73.06 ± 0.50	73.53 ± 0.51
	0.1	0.2	73.02 ± 0.50	72.68 ± 0.16	72.31 ± 0.25	72.88 ± 0.14	71.92 ± 1.98	73.81 ± 0.40
	0.1	0.3	72.44 ± 0.47	72.15 ± 0.23	69.06 ± 2.01	72.26 ± 0.43	69.53 ± 4.90	73.34 ± 1.27
	0.1	0.4	54.87 ± 0.85	71.48 ± 0.50	63.60 ± 1.04	71.44 ± 0.72	72.43 ± 1.90	73.56 ± 0.89
	0.2	0.3	72.81 ± 0.51	72.43 ± 0.14	71.44 ± 0.93	72.78 ± 0.28	71.55 ± 2.04	73.75 ± 0.26
	0.2	0.4	72.06 ± 0.19	71.97 ± 0.41	63.49 ± 1.58	71.97 ± 0.37	65.99 ± 7.99	71.43 ± 2.26
Compas	0.0	0.1	66.36 ± 1.05	66.04 ± 1.14	66.16 ± 1.13	68.06 ± 0.70	67.14 ± 0.92	68.22 ± 0.68
$n = 7,168$	0.0	0.2	66.84 ± 0.69	66.06 ± 0.81	65.38 ± 1.40	68.03 ± 0.77	66.51 ± 1.90	68.40 ± 0.78
$d = 10$	0.0	0.3	58.06 ± 0.32	62.69 ± 1.20	53.04 ± 3.69	66.41 ± 1.19	59.02 ± 7.78	65.93 ± 0.56
	0.0	0.4	51.16 ± 0.30	62.41 ± 0.71	54.03 ± 4.95	65.01 ± 0.65	54.26 ± 2.50	65.85 ± 1.17
	0.1	0.2	66.41 ± 0.43	65.69 ± 0.57	65.91 ± 0.97	67.49 ± 0.40	66.54 ± 0.21	67.80 ± 0.44
	0.1	0.3	65.91 ± 0.42	65.22 ± 0.63	61.24 ± 0.70	67.36 ± 0.79	65.76 ± 2.09	68.05 ± 0.56
	0.1	0.4	51.60 ± 0.12	63.34 ± 1.12	57.65 ± 3.90	66.47 ± 1.34	55.83 ± 6.43	67.04 ± 0.75
	0.2	0.3	65.06 ± 0.72	65.86 ± 1.69	65.06 ± 1.48	68.02 ± 0.94	66.46 ± 1.27	68.04 ± 1.11
	0.2	0.4	64.82 ± 0.52	65.47 ± 0.46	59.68 ± 2.49	67.37 ± 0.54	63.85 ± 3.31	68.39 ± 0.56
Fairface	0.0	0.1	87.64 ± 0.03	87.75 ± 0.03	87.41 ± 0.11	87.58 ± 0.15	88.23 ± 0.07	88.49 ± 0.12
$n = 108,501$	0.0	0.2	85.22 ± 0.06	85.83 ± 0.08	85.08 ± 0.16	85.18 ± 0.16	88.55 ± 0.03	88.67 ± 0.03
$d = 50$	0.0	0.3	81.51 ± 0.09	83.36 ± 0.04	79.62 ± 0.12	81.37 ± 0.35	87.44 ± 0.15	88.25 ± 0.06
	0.1	0.2	87.67 ± 0.07	87.56 ± 0.04	87.21 ± 0.08	87.28 ± 0.05	88.45 ± 0.06	88.65 ± 0.07
	0.1	0.3	72.03 ± 0.13	85.68 ± 0.07	83.20 ± 0.12	84.58 ± 0.09	87.81 ± 0.14	88.50 ± 0.12
	0.1	0.4	59.30 ± 0.11	83.10 ± 0.08	74.56 ± 0.53	80.51 ± 0.30	80.83 ± 2.24	87.10 ± 0.39
	0.2	0.3	74.18 ± 0.20	87.34 ± 0.14	86.47 ± 0.09	87.00 ± 0.11	88.46 ± 0.08	88.58 ± 0.10
	0.2	0.4	58.30 ± 0.23	85.48 ± 0.09	78.33 ± 0.63	84.05 ± 0.13	81.90 ± 0.58	87.69 ± 0.15

Table B2: **Accuracy of compared methods across different levels of noise gap for multi-class classification.** Mis. SL: surrogate loss [25] with misspecified parameters. Est. SL: surrogate loss [25] with estimated parameters. CE: vanilla cross entropy. Peer: peer loss function [21]. When noise gap is less than 0.2, cross entropy with increasing-to-balancing reaches a higher accuracy than cross entropy at a lower noise. When noise gap is 0.3, balancing cannot compensate for the loss of increasing noise.

		BASELINES (LESS NOISE)				NOISE+ (MORE NOISE)	
Dataset	noise gap	Mis. SL	Est. SL	CE	Peer	CE	Peer
MNIST	0.1	89.59 ± 0.01	89.69 ± 0.07	86.66 ± 0.54	88.12 ± 0.01	86.81 ± 0.62	89.19 ± 0.05
	0.2	88.10 ± 0.10	88.61 ± 0.16	84.53 ± 1.60	87.21 ± 0.53	85.97 ± 0.69	89.12 ± 0.24
	0.3	84.97 ± 0.11	86.88 ± 0.17	85.24 ± 1.05	86.35 ± 0.33	81.89 ± 1.54	88.75 ± 0.19
CIFAR-10	0.1	70.90 ± 2.66	85.76 ± 1.44	88.03 ± 1.07	89.66 ± 1.18	88.69 ± 0.82	89.90 ± 0.52
	0.2	80.51 ± 4.51	86.34 ± 2.30	88.43 ± 1.29	89.36 ± 0.56	89.01 ± 1.27	90.08 ± 1.26
	0.3	81.30 ± 2.31	90.61 ± 0.52	89.78 ± 1.16	90.24 ± 1.05	87.98 ± 1.29	89.92 ± 0.92

Table B3: **Constrained learning results with group-dependent label noise.** LR: naïve logistic regression without noise correction. GPR: group-weighted peer loss [30]. Peer: peer loss [21].

				LESS NOISE		MORE NOISE	
Dataset	e_a	e_b	Metrics	LR	GPL	LR	Peer
Adult	0.1	0.3	*accuracy*	72.57	71.92	71.07	73.21
			fairness	2.37	3.39	1.83	1.95
	0.2	0.3	*accuracy*	72.4	72.92	73.07	71.8
			fairness	6.67	3.36	4.21	0.93
	0.2	0.4	*accuracy*	72.73	71.2	71.88	73.02
			fairness	6.48	2.95	3.16	1.67
	0.3	0.4	*accuracy*	73.15	73.74	71.36	72.74
			fairness	5.29	4.11	5.49	1.88
Compas	0.1	0.3	*accuracy*	63.88	63.73	64.56	64.33
			fairness	7.17	6.58	7.35	1.89
	0.2	0.3	*accuracy*	63.73	63.28	64.26	67.8
			fairness	10.52	4.47	7.10	2.76
	0.2	0.4	*accuracy*	62.60	66.03	66.22	64.15
			fairness	2.87	7.55	6.07	3.63
	0.3	0.4	*accuracy*	61.93	62.08	61.63	62.68
			fairness	17.97	3.06	7.70	3.74
Fairface	0.2	0.4	*accuracy*	86.97	87.47	88.19	87.93
			fairness	5.87	4.70	1.38	0.25
	0.1	0.3	*accuracy*	88.23	88.23	88.58	88.60
			fairness	5.53	4.93	2.11	2.17
	0.0	0.2	*accuracy*	88.61	88.53	88.90	88.85
			fairness	4.05	3.75	2.64	2.20
	0.0	0.1	*accuracy*	89.08	88.84	89.00	89.05
			fairness	3.99	3.92	2.97	2.91
	0.2	0.3	*accuracy*	88.63	88.78	88.80	88.83
			fairness	3.50	3.14	2.19	1.33