Error-Bounded Correction of Noisy Labels
— Supplementary Material —

Songzhu Zheng [1] **Pengxiang Wu** [2] **Aman Goswami** [3] **Mayank Goswami** [4] **Dimitris Metaxas** [2] **Chao Chen** [5]

1. Additional (Synthetic) Experiment for Validation of the Bound

In Section 2.3 of the submitted manuscript, we used the output of deep neural networks f as an approximation of η on the CIFAR10 dataset. We provided empirical estimates of the constants C and λ in the Tsybakov condition for η, as well as estimates of the probability $\Pr[\widetilde{y} = h^*(\boldsymbol{x}), f_{\widetilde{y}}(\boldsymbol{x}) < \Delta]$.

In this section, we provide additional experiments on a *synthetic data set* generated using a mixture-of-Gaussians distribution. In this ideal setting, we know η, τ_{01}, τ_{10}, $\widetilde{\eta}$ *exactly*. We can a) use $\widetilde{\eta}$ as the classifier and b) evaluate the constants in Tsybakov condition for η in order to evaluate the upper bound in Theorem 1.

Estimation of Tsybakov condition constants. We let $\Pr(\boldsymbol{x})$ be a mixture of Gaussian distribution in a 10 dimensional feature space, $\boldsymbol{x} \sim \frac{1}{2}\mathcal{N}(0, I_{10\times 10}) + \frac{1}{2}\mathcal{N}(1, I_{10\times 10})$. We sample from the two components with equal probability. If $\boldsymbol{x}$ comes from component $\mathcal{N}(0, I_{10\times 10})$, it is given label 0. Otherwise, if $\boldsymbol{x}$ comes from component $\mathcal{N}(1, I_{10\times 10})$, it is given label 1. The true conditional distribution is $\eta(\boldsymbol{x}) = \frac{\exp\left\{-\frac{1}{2}||\boldsymbol{x}-1||^2\right\}}{\exp\left\{-\frac{1}{2}||\boldsymbol{x}||^2\right\}+\exp\left\{-\frac{1}{2}||\boldsymbol{x}-1||^2\right\}}$.

Following the idea of our experiment on CIFAR10 in the manuscript (Section 2.4), we estimate $\Pr\left[|\eta(\boldsymbol{x}) - \frac{1}{2}| \le t\right]$ for values of t sampled between 0 and 0.9 using the empirical frequency $p_t = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{|\eta(\boldsymbol{x})-1/2| \le t\}}(\boldsymbol{x})$. Note that if the Tsybakov condition is tight, $\log(p_t)$ approximates $\log(Ct^\lambda)$. The samples for $\log(t)$ and correspondingly, $\log(Ct^\lambda) \approx \log(p_t)$ are drawn as blue dots in Figure 1(a). The ordinary least square (OLS) linear regression results is drawn as a red line. We found the estimated values of C and λ to be 0.58 and 1.27 respectively. The estimation is high is confidence: the determinant coefficient R^2 equals 0.904, and we have a p-value which is less than 10^{-4}.

Estimation of the error bound, and its tightness. We also introduce label noise using predefined transition probability τ_{01} and τ_{10}. We can estimate C and λ as mentioned above, and know $\tau_{01}, \tau_{10}, \eta(x)$, and thus, $\widetilde{\eta}(\boldsymbol{x})$. Therefore we can evaluate the error bound in Theorem 1. We plot the error bound as a function of ϵ in Figures 1(b) and (c) (drawn green curves).

Finally, we assume a perfect noisy classifier $f = \widetilde{\eta}$. In other words, $\epsilon = 0$. We empirically show that when $f(\boldsymbol{x}) < \Delta$, the probability of $\widetilde{y}$ being correct (i.e., $\widetilde{y} = h^*(\boldsymbol{x})$) is zero (blue lines in Figures 1(b) and (c)).

Validation of the label-correction algorithm. To the same synthetic dataset, we also apply our LRT-Correction algorithm and validate the bound in Corollary 1. Since we know $\widetilde{\eta}(\boldsymbol{x})$, τ_{01} and τ_{10}, we calculate the correction error bound of Corollary 1 in closed form. We draw the bound w.r.t. the error ϵ in orange curves in Figure 2. Finally, we run our label correction algorithm using the perfect noisy classifier $f = \widetilde{\eta}$ and validate that the corrected labels are very close to clean (the success rate is limited by the asymmetry level of the noise pattern). See blue lines in Figure 2.

[1]Department of Applied Mathematics and Statistics, Stony Brook University, NY, USA [2]Department of Computer Science, Rutgers University, NJ, USA [3]Bain & Company, Bangalore, India. [4]Department of Computer Science, City University of New York, NY, USA [5]Department of Biomedical Informatics, Stony Brook University, NY, USA. Correspondence to: Songzhu Zheng <zheng.songzhu@stonybrook.edu>.

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 119, 2020.

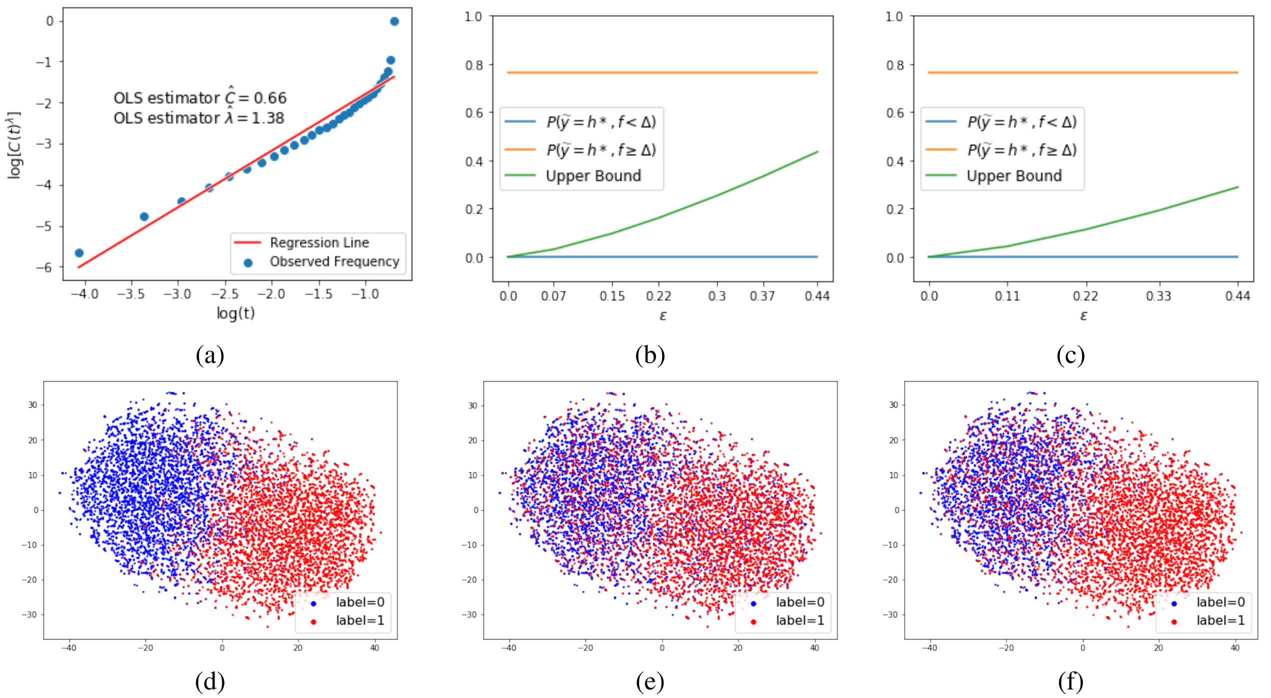


Figure 1. Synthetic experiment using Mixture of Gaussian at noise level 20%. (a): Check of Tsybakov condition using linear regression, where y-axis is the proportion of data points at distance t from decision boundary. (b): Proportion of labels that are not correct (not consistent with Bayes optimal decision rule) and the proposed upper bound. (c): Same as (b) but labels are corrupted with aysmmetric noise. (d): t-SNE of the clean data. (e): t-SNE of the data with symmetric noise. (f): t-SNE of the data with asymmetric noise.

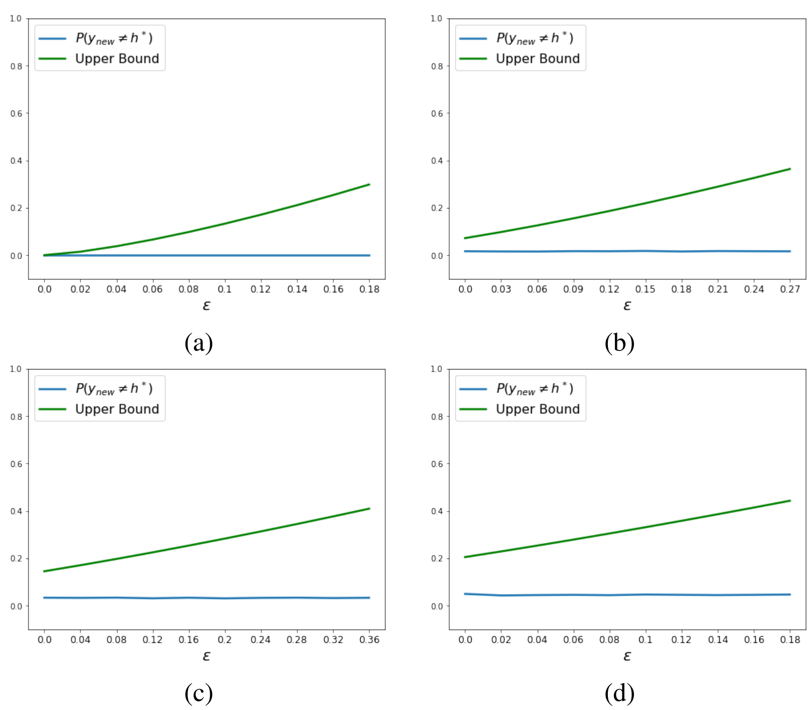
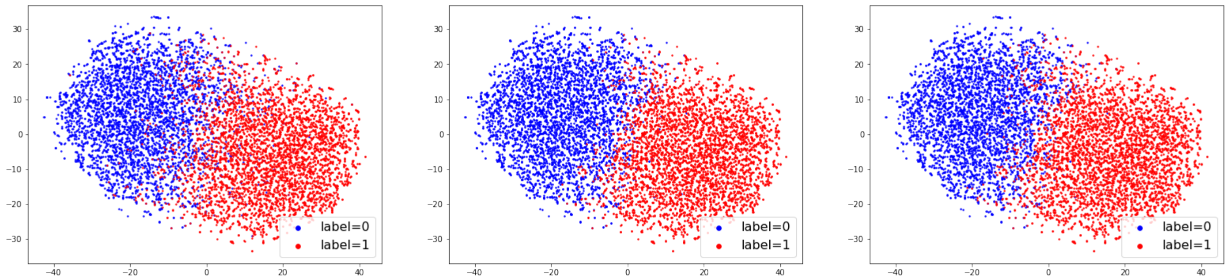


Figure 2. Performance of LRT algorithm given $\widetilde{\eta}(\boldsymbol{x})$ v.s the proposed upper bound. (a): Symmetric noise ($\tau_{10} = \tau_{01} = 0.3$). (b): Asymmetric noise ($\tau_{10} = 0.2, \tau_{01} = 0.3$). (c): Asymmetric noise ($\tau_{10} = 0.1, \tau_{01} = 0.3$). (d): Asymmetric noise ($\tau_{10} = 0.3, \tau_{01} = 0$)



(a) (b) (c)

Figure 3. Label Correction Result Using LRT-Correct. (a): Clean data as it in Fig 1d. (b): Labels after correction for data in Fig 1e. (c): Labels after correction for data in Fig 1f.

2. Proof of Theorem 2

For convenience, we restate some notation here. Let $\eta_i(\boldsymbol{x} \mid i, j) = \Pr(y = i \mid \boldsymbol{x}, y \in \{i, j\})$, $\widetilde{\eta}_i(\boldsymbol{x} \mid i, j) = \Pr(\widetilde{y} = i \mid \boldsymbol{x}, y \in \{i, j\})$. Note whenever we condition on pair of i, j, we condition on the true label class $y \in \{i, j\}$. Let $f_i(\boldsymbol{x}|i, j) := \frac{f_i(\boldsymbol{x})}{\eta_i(\boldsymbol{x})+\eta_j(\boldsymbol{x})}$. With this definition, we have $\frac{f_i(\boldsymbol{x}|i,j)}{f_j(\boldsymbol{x}|i,j)} = \frac{f_i(\boldsymbol{x})}{f_j(\boldsymbol{x})}$. Also define $m_{\boldsymbol{x}} := \arg\max_i f_i(\boldsymbol{x})$, $u_{\boldsymbol{x}} := \arg\max_i \eta_i(\boldsymbol{x})$ and $s_{\boldsymbol{x}} := \arg\max_{i \neq u_{\boldsymbol{x}}} \eta_i(\boldsymbol{x})$.

For any pair of labels i, j, we have the linear relationship (remember the noise is independent from the feature):

$$
\begin{aligned}
\widetilde{\eta}_i(\boldsymbol{x}|i, j) &= \Pr(\widetilde{y} = i \mid \boldsymbol{x}, y \in \{i, j\}) = \Pr(\widetilde{y} = i, y = i \mid \boldsymbol{x}, y \in \{i, j\}) + \Pr(\widetilde{y} = i, y = j \mid \boldsymbol{x}, y \in \{i, j\}) \\
&= \Pr(\widetilde{y} = i \mid y = i) \Pr(y = i \mid \boldsymbol{x}, y \in \{i, j\}) + \Pr(\widetilde{y} = i \mid y = j) \Pr(y = j \mid \boldsymbol{x}, y \in \{i, j\}) \\
&= \tau_{ii}\eta_i(\boldsymbol{x} \mid i, j) + \tau_{ji}\eta_j(\boldsymbol{x} \mid i, j) \\
&= (\tau_{ii} - \tau_{ji})\eta_i(\boldsymbol{x} \mid i, j) + \tau_{ji}
\end{aligned}
$$

Finally define $[Nc] := \{1, 2, \cdots, N_c\}$.

We also restate the multi-class Tsybakov condition here:

Assumption 1 (Multi-class Tsybakov Condition)**.** *$\exists C, \lambda > 0$ and $t_0 \in (0, 1]$ such that for all $t \le t_0$,*

$$
\Pr\left[|\eta_{u_{\boldsymbol{x}}}(\boldsymbol{x}) - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})| \le t\right] \le Ct^{\lambda}
$$

Define $\epsilon_{i,j}(\boldsymbol{x}) := \left| f_i(\boldsymbol{x} \mid i, j) - \frac{\widetilde{\eta}_i(\boldsymbol{x})}{\eta_i(\boldsymbol{x})+\eta_j(\boldsymbol{x})} \right|$ and $\epsilon := \max_{\boldsymbol{x}} |\epsilon_{\widetilde{y}, m_{\boldsymbol{x}}}(\boldsymbol{x})|$.

Observe following:

$$
\frac{\widetilde{\eta}_i(\boldsymbol{x})}{\eta_i(\boldsymbol{x}) + \eta_{m_{\boldsymbol{x}}}(\boldsymbol{x})} = \frac{\sum_k \tau_{ki}\eta_k(\boldsymbol{x})}{\eta_i(\boldsymbol{x}) + \eta_{m_{\boldsymbol{x}}}(\boldsymbol{x})} \ge \frac{\tau_{ii}\eta_i(\boldsymbol{x}) + \tau_{m_{\boldsymbol{x}},i}\eta_{m_{\boldsymbol{x}}}(\boldsymbol{x})}{\eta_i(\boldsymbol{x}) + \eta_{m_{\boldsymbol{x}}}(\boldsymbol{x})} = \widetilde{\eta}_i(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})
$$

For an input feature $\boldsymbol{x}$ and label $\widetilde{y} = i$, Let $\delta_0(\boldsymbol{x}) = \frac{(\tau_{i,i} - \tau_{m_{\boldsymbol{x}},i})\eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) + \tau_{m_{\boldsymbol{x}},i}}{f_{m_{\boldsymbol{x}}}(\boldsymbol{x}|i, m_{\boldsymbol{x}})}$, $\delta_1 = \min(1, \min_{\boldsymbol{x}} \delta_0(\boldsymbol{x}))$. $\forall \delta_0 > 0$, $C_1 := \frac{\delta_0 f_{m_{\boldsymbol{x}}}(\boldsymbol{x}|i, m_{\boldsymbol{x}}) - \tau_{m_{\boldsymbol{x}},i}}{(\tau_{i,i} - \tau_{m_{\boldsymbol{x}},i})} - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})$. Finally, let $\Delta = \delta_1 \min_{\boldsymbol{x}} f_{m_{\boldsymbol{x}}}(\boldsymbol{x})$.

Theorem 2. *Assume $\eta(\boldsymbol{x})$ fulfills multi-class Tsybakov condition for constant $C, \lambda > 0$ and $t_0 \in (0, \frac{1}{2}]$. Assume that $\epsilon \le \frac{t_0(\tau_{i,i} - \tau_{j,i})}{\delta_0 + 1}$ $\forall i, j \in [N_c]$. Then for $\Delta = \min_{\boldsymbol{x}} \frac{(\tau_{\widetilde{y},\widetilde{y}} - \tau_{m_{\boldsymbol{x}},\widetilde{y}})\eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) + \tau_{m_{\boldsymbol{x}},\widetilde{y}}}{f_{m_{\boldsymbol{x}}}(\boldsymbol{x}|\widetilde{y}, m_{\boldsymbol{x}})} \min_{\boldsymbol{x}} f_{m_{\boldsymbol{x}}}(\boldsymbol{x}) > 0$:*

$$
\Pr_{(x,y)\sim D}\left[\widetilde{y} = h^*(\boldsymbol{x}), f_{\widetilde{y}}(\boldsymbol{x}) < \Delta\right] \le C\left[O(\epsilon)\right]^{\lambda}
$$

Proof. Unlike the binary case where the decision boundary is determined by a constant, in the multi-class case, the decision boundary is determined by the second largest element of $\eta(\boldsymbol{x})$, i.e. $\eta_s(\boldsymbol{x})$. The margin we want to bound is the difference between $\eta_m(\boldsymbol{x})$ and $\eta_s(\boldsymbol{x})$.

Case 1: when $f_i(\boldsymbol{x}) \neq f_{m_{\boldsymbol{x}}}(\boldsymbol{x})$.

$$
\begin{aligned}
&\Pr[\widetilde{y} = h^*(\boldsymbol{x}), f_i(\boldsymbol{x})/f_{m_{\boldsymbol{x}}}(\boldsymbol{x}) < \delta_0] = \Pr[\widetilde{y} = h^*(\boldsymbol{x}), f_i(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) < \delta_0 f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})] \\
&\leq \Pr\left[\widetilde{y} = h^*(\boldsymbol{x}), \frac{\widetilde{\eta}_i(\boldsymbol{x})}{\eta_i(\boldsymbol{x}) + \eta_{m_{\boldsymbol{x}}}(\boldsymbol{x})} - \epsilon < \delta_0 f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})\right] \\
&\leq \Pr[\widetilde{y} = h^*(\boldsymbol{x}), \widetilde{\eta}_i(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) - \epsilon < \delta_0 f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})] \\
&= \Pr\left[\eta_i(\boldsymbol{x}) > \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}), \eta_i(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) < \frac{\delta_0 f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) - \tau_{m_{\boldsymbol{x}},i}}{\tau_{ii} - \tau_{m_{\boldsymbol{x}},i}} + \frac{\epsilon}{\tau_{ii} - \tau_{m_{\boldsymbol{x}},i}}\right] \\
&\leq \Pr\left[\eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) < \eta_i(\boldsymbol{x}) < \frac{\delta_0 f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) - \tau_{m_{\boldsymbol{x}},i}}{(\tau_{ii} - \tau_{m_{\boldsymbol{x}},i})} + \frac{\epsilon}{(\tau_{ii} - \tau_{m_{\boldsymbol{x}},i})}\right] \\
&= \Pr\left[0 < \eta_i(\boldsymbol{x}) - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) < \frac{\delta_0 f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) - \tau_{m_{\boldsymbol{x}},i}}{(\tau_{ii} - \tau_{m_{\boldsymbol{x}},i})} - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) + \frac{\epsilon}{(\tau_{ii} - \tau_{m_{\boldsymbol{x}},i})}\right] \\
&\leq C\left[\frac{\delta_0 f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) - \tau_{m_{\boldsymbol{x}},i}}{(\tau_{ii} - \tau_{m_{\boldsymbol{x}},i})} - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) + \frac{\epsilon}{(\tau_{ii} - \tau_{m_{\boldsymbol{x}},i})}\right]^\lambda = C\left[C_1 + O(\epsilon)\right]^\lambda
\end{aligned}
$$

Notice that to derive the third inequality, we use the simple fact that $\eta_i(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) = \frac{P(y=i|\boldsymbol{x})}{P(y\in\{i,m_{\boldsymbol{x}}\}|\boldsymbol{x})} \geq P(y = i \mid \boldsymbol{x}) = \eta_i(\boldsymbol{x})$. The final inequality is gained by using the multi-class Tsybakov condition. Notice that if we plug in $\delta_0(\boldsymbol{x}) = \frac{(\tau_{ii}-\tau_{m_{\boldsymbol{x}},i})\eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})+\tau_{m_{\boldsymbol{x}},i}}{f_{m_{\boldsymbol{x}}}(\boldsymbol{x}|i,m_{\boldsymbol{x}})} > 0$, then we get $C_1 = \frac{\delta_0 f_{m_{\boldsymbol{x}}}(\boldsymbol{x}|i,m_{\boldsymbol{x}})-\tau_{m_{\boldsymbol{x}},i}}{(\tau_{ii}-\tau_{m_{\boldsymbol{x}},i})} - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) = 0$. This implies that if we pick $0 < \delta_0(\boldsymbol{x}) < \frac{(\tau_{ii}-\tau_{m_{\boldsymbol{x}},i})\eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})+\tau_{m_{\boldsymbol{x}},i}}{f_{m_{\boldsymbol{x}}}(\boldsymbol{x}|i,m_{\boldsymbol{x}})}$, then $C_1 < 0$.

Now since $\delta_1 = \min\left(1, \min_{\boldsymbol{x}} \delta_0(\boldsymbol{x})\right)$ and $\Delta = \delta_1 \min_{\boldsymbol{x}} f_{m_{\boldsymbol{x}}}(\boldsymbol{x})$, then:

$$
\Pr\left[\widetilde{y} = h^*(\boldsymbol{x}), f_i(\boldsymbol{x}) < \Delta\right] = \Pr\left[\widetilde{y} = h^*(\boldsymbol{x}), f_i(\boldsymbol{x})/\min_{\boldsymbol{x}} f_{m_{\boldsymbol{x}}} < \delta_1\right] \leq \Pr[\widetilde{y} = h^*(\boldsymbol{x}), f_i(\boldsymbol{x})/f_{m_{\boldsymbol{x}}}(\boldsymbol{x}) < \delta_0(\boldsymbol{x})] \leq C\left[O(\epsilon)\right]^\lambda
$$

completes the proof of the theorem in this case.

Case 2: when $f_i(\boldsymbol{x}) = f_{m_{\boldsymbol{x}}}(\boldsymbol{x})$.:

$$
\Pr[\widetilde{y} = h^*(\boldsymbol{x}), f_i(\boldsymbol{x}) < \Delta, f_i(\boldsymbol{x}) = f_{m_{\boldsymbol{x}}}(\boldsymbol{x})] = \Pr[\widetilde{y} = h^*(\boldsymbol{x}), f_{m_{\boldsymbol{x}}}(\boldsymbol{x}) < \Delta] = 0
$$

Case1 and Case2 together complete the proof. □

Lemma 1. *(Algorithm Multiclass-Theorem Guarantee). $\forall i, j \in [N_c]$, assume $\eta(\boldsymbol{x})$ fulfills multi-class Tsybakov condition for constant $C > 0$, $\lambda > 0$ and $t_0 \in (0, 1]$. Assume that $\epsilon \leq \frac{t_0 \min(\tau_{ii}-\tau_{j,i})}{1/\delta+1}$, $\forall i, j \in [N_c]$. Let $\widetilde{y}_{new}$ denotes the output of the* `LRT-Correction` *with $\boldsymbol{x}$, $\widetilde{y}_{\boldsymbol{x}}$, f, and the given δ, then for the constant $C_1 = \max_{\boldsymbol{x}} \frac{1-\tau_{m_{\boldsymbol{x}},u_{\boldsymbol{x}}}}{(\tau_{u_{\boldsymbol{x}},u_{\boldsymbol{x}}}-\tau_{m_{\boldsymbol{x}},u_{\boldsymbol{x}}})} - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})$:*

1. *Sensitivity Optimized Critical Value. Let $\delta = \min_{\boldsymbol{x}} \frac{(\tau_{\widetilde{y},\widetilde{y}}-\tau_{m_{\boldsymbol{x}},\widetilde{y}})\eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})+\tau_{m_{\boldsymbol{x}},\widetilde{y}}}{f_{m_{\boldsymbol{x}}}(\boldsymbol{x}|\widetilde{y},m_{\boldsymbol{x}})}$ and :*

$$
\Pr_{(x,y)\sim D}\left[\widetilde{y}_{new} \neq h^*(\boldsymbol{x}) \mid \widetilde{y} \text{ is rejected}\right] \leq C \max\left[O(\epsilon), C_1\right]^\lambda
$$

2. *Specificity Optimized Critical Value. Let $\delta = \max_{\boldsymbol{x}} \frac{f_{\widetilde{y}}(\boldsymbol{x}|m_{\boldsymbol{x}},\widetilde{y})}{(\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}}-\tau_{\widetilde{y},m_{\boldsymbol{x}}})\eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})+\tau_{\widetilde{y},m_{\boldsymbol{x}}}}$ and :*

$$
\Pr_{(x,y)\sim D}\left[\widetilde{y}_{new} \neq h^*(\boldsymbol{x}) \mid \widetilde{y} \text{ is accepted}\right] \leq C \max\left[O(\epsilon), C_1\right]^\lambda
$$

Proof. For a input $(\boldsymbol{x}, \widetilde{y})$, if $\widetilde{y}$ is rejected, we want to bound the probability of event $\{\eta_i(\boldsymbol{x}) = \max \eta(\boldsymbol{x})\}$ and if $\widetilde{y}$ is accepted, we want to bound the probability of event $\{\eta_{m_{\boldsymbol{x}}}(\boldsymbol{x}) = \max \eta(\boldsymbol{x})\}$. Note $m_{\boldsymbol{x}}$ here is the class that maximize $f(\boldsymbol{x})$ but not necessarily a maximizer of $\eta(\boldsymbol{x})$, i.e. $u_{\boldsymbol{x}}$. All these events could be summarized into three cases (the assumption on ϵ will guarantee the Tsybakov condition):

Case 1: $\widetilde{y} = i$ is rejected and $\eta_i(\boldsymbol{x}) = \max \eta(\boldsymbol{x}) = \eta_{u_{\boldsymbol{x}}}(\boldsymbol{x})$:

$$\begin{aligned}
&\Pr[\widetilde{y} = h^*(\boldsymbol{x}), f_i(\boldsymbol{x})/f_{m_{\boldsymbol{x}}}(\boldsymbol{x}) < \delta] = \Pr[\widetilde{y} = h^*(\boldsymbol{x}), f_i(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) < \delta f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})] \\
&\leq \Pr\left[\widetilde{y} = h^*(\boldsymbol{x}), \frac{\widetilde{\eta}_i(\boldsymbol{x})}{\eta_i(\boldsymbol{x}) + \eta_{m_{\boldsymbol{x}}}(\boldsymbol{x})} - \epsilon < \delta_0 f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})\right] \\
&\leq \Pr\left[\widetilde{y} = h^*(\boldsymbol{x}), \widetilde{\eta}_i(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) - \epsilon < \delta f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})\right] \\
&= \Pr\left[\eta_i(\boldsymbol{x}) > \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}), \eta_i(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) < \frac{\delta f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) - \tau_{m_{\boldsymbol{x}},i}}{\tau_{ii} - \tau_{i,m_{\boldsymbol{x}}}} + \frac{\epsilon}{\tau_{ii} - \tau_{i,m_{\boldsymbol{x}}}}\right] \\
&\leq \Pr\left[\eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) < \eta_i(\boldsymbol{x}) < \frac{\delta f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) - \tau_{m_{\boldsymbol{x}},i}}{(\tau_{ii} - \tau_{i,m_{\boldsymbol{x}}})} + \frac{\epsilon}{(\tau_{ii} - \tau_{i,m_{\boldsymbol{x}}})}\right] \\
&= \Pr\left[0 < \eta_i(\boldsymbol{x}) - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) < \frac{\delta f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) - \tau_{m_{\boldsymbol{x}},i}}{(\tau_{ii} - \tau_{i,m_{\boldsymbol{x}}})} - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) + \frac{\epsilon}{(\tau_{ii} - \tau_{i,m_{\boldsymbol{x}}})}\right] \\
&\leq C\left[\frac{\delta f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) - \tau_{m_{\boldsymbol{x}},i}}{(\tau_{ii} - \tau_{i,m_{\boldsymbol{x}}})} - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) + \frac{\epsilon}{(\tau_{ii} - \tau_{i,m_{\boldsymbol{x}}})}\right]^\lambda \qquad (1)
\end{aligned}$$

Case 2: $\widetilde{y} = i$ is accepted but $\eta_{m_{\boldsymbol{x}}} = \max \eta(\boldsymbol{x}) = \eta_{u_{\boldsymbol{x}}}(\boldsymbol{x})$:

$$\begin{aligned}
&\Pr\left[\frac{f_i(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})}{f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})} \geq \delta, \eta_{m_{\boldsymbol{x}}}(\boldsymbol{x}) = \max \eta(\boldsymbol{x})\right] = \Pr\left[f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) \leq f_i(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})/\delta, \eta_{m_{\boldsymbol{x}}}(\boldsymbol{x}) \geq \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})\right] \\
&\leq \Pr\left[\eta_{m_{\boldsymbol{x}}}(\boldsymbol{x}) > \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}), \frac{\widetilde{\eta}_{m_{\boldsymbol{x}}}(\boldsymbol{x})}{\eta_i(\boldsymbol{x}) + \eta_{m_{\boldsymbol{x}}}(\boldsymbol{x})} - \epsilon < f_i(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})/\delta\right] \\
&\leq \Pr\left[\eta_{m_{\boldsymbol{x}}}(\boldsymbol{x}) > \eta_s(\boldsymbol{x}), \widetilde{\eta}_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) - \epsilon \leq f_i(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})/\delta\right] \\
&= \Pr\left[\eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) < \eta_{m_{\boldsymbol{x}}}(\boldsymbol{x}) < \eta_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid m_{\boldsymbol{x}}, i) \leq \frac{f_i(\boldsymbol{x} \mid m_{\boldsymbol{x}}, i)/\delta - \tau_{i,m_{\boldsymbol{x}}}}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}} - \tau_{i,m_{\boldsymbol{x}}}} + \frac{\epsilon}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}} - \tau_{i,m_{\boldsymbol{x}}}}\right] \\
&= \Pr\left[0 < \eta_{m_{\boldsymbol{x}}}(\boldsymbol{x}) - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) < \frac{f_i(\boldsymbol{x} \mid m_{\boldsymbol{x}}, i)/\delta - \tau_{i,m_{\boldsymbol{x}}}}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}} - \tau_{i,m_{\boldsymbol{x}}}} - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) + \frac{\epsilon}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}} - \tau_{i,m_{\boldsymbol{x}}}}\right] \\
&\leq C\left[\frac{f_i(\boldsymbol{x} \mid m_{\boldsymbol{x}}, i)/\delta - \tau_{i,m_{\boldsymbol{x}}}}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}} - \tau_{i,m_{\boldsymbol{x}}}} - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) + \frac{\epsilon}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}} - \tau_{i,m_{\boldsymbol{x}}}}\right]^\lambda \qquad (2)
\end{aligned}$$

Case 3: $u_{\boldsymbol{x}} \neq i$ and $u_{\boldsymbol{x}} \neq m_{\boldsymbol{x}}$, no matter $\widetilde{y}$ is rejected or not.

$$\begin{aligned}
&\Pr\left[u_{\boldsymbol{x}} \neq i, u_{\boldsymbol{x}} \neq m_{\boldsymbol{x}}\right] = \Pr\left[f_{u_{\boldsymbol{x}}}(\boldsymbol{x}) < f_{m_{\boldsymbol{x}}}(\boldsymbol{x}), \eta_{u_{\boldsymbol{x}}}(\boldsymbol{x}) > \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})\right] \\
&\leq \Pr\left[f_{u_{\boldsymbol{x}}}(\boldsymbol{x} \mid u_{\boldsymbol{x}}, m_{\boldsymbol{x}}) < 1, \eta_{u_{\boldsymbol{x}}}(\boldsymbol{x}) > \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})\right] \\
&\leq \Pr\left[\frac{\widetilde{\eta}_{u_{\boldsymbol{x}}}(\boldsymbol{x})}{\eta_{u_{\boldsymbol{x}}}(\boldsymbol{x}) + \eta_{m_{\boldsymbol{x}}}(\boldsymbol{x})} - \epsilon < 1, \eta_{u_{\boldsymbol{x}}}(\boldsymbol{x}) > \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})\right] \\
&\leq \Pr\left[\widetilde{\eta}_{u_{\boldsymbol{x}}}(\boldsymbol{x} \mid u_{\boldsymbol{x}}, m_{\boldsymbol{x}}) - \epsilon < 1, \eta_{u_{\boldsymbol{x}}}(\boldsymbol{x}) > \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})\right] \\
&= \Pr\left[\widetilde{\eta}_{u_{\boldsymbol{x}}}(\boldsymbol{x} \mid u_{\boldsymbol{x}}, m_{\boldsymbol{x}}) \leq 1, \eta_{u_{\boldsymbol{x}}}(\boldsymbol{x}) > \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})\right] \\
&\leq \Pr\left[\eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) \leq \eta_{u_{\boldsymbol{x}}}(\boldsymbol{x}) \leq \eta_{u_{\boldsymbol{x}}}(\boldsymbol{x} \mid u_{\boldsymbol{x}}, m_{\boldsymbol{x}}) \leq \frac{1 - \tau_{m_{\boldsymbol{x}},u_{\boldsymbol{x}}}}{(\tau_{u_{\boldsymbol{x}},u_{\boldsymbol{x}}} - \tau_{m_{\boldsymbol{x}},u_{\boldsymbol{x}}})}\right] \\
&= \Pr\left[0 \leq \eta_{u_{\boldsymbol{x}}}(\boldsymbol{x}) - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) \leq \frac{1 - \tau_{m_{\boldsymbol{x}},u_{\boldsymbol{x}}}}{(\tau_{u_{\boldsymbol{x}},u_{\boldsymbol{x}}} - \tau_{m_{\boldsymbol{x}},u_{\boldsymbol{x}}})} - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})\right] \\
&\leq C\left[\frac{1 - \tau_{m_{\boldsymbol{x}},u_{\boldsymbol{x}}}}{(\tau_{u_{\boldsymbol{x}},u_{\boldsymbol{x}}} - \tau_{m_{\boldsymbol{x}},u_{\boldsymbol{x}}})} - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})\right]^\lambda \qquad (3)
\end{aligned}$$

Case 1 and case 2 can be controlled by the choice of δ.

For case 1, if we set $\delta = \min_{\boldsymbol{x}} \frac{(\tau_{\widetilde{y},\widetilde{y}} - \tau_{m_{\boldsymbol{x}},\widetilde{y}})\eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) + \tau_{m_{\boldsymbol{x}},\widetilde{y}}}{f_{m_{\boldsymbol{x}}}(\boldsymbol{x}|\widetilde{y}, m_{\boldsymbol{x}})}$, then by algebra $\frac{\delta f_{m_{\boldsymbol{x}}}(\boldsymbol{x}|\widetilde{y}, m_{\boldsymbol{x}}) - \tau_{m_{\boldsymbol{x}},\widetilde{y}}}{(\tau_{\widetilde{y},\widetilde{y}} - \tau_{\widetilde{y},m_{\boldsymbol{x}}})} - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) < 0$, which implies that (1) is upper bounded by $C[O(\epsilon)]^{\lambda}$. Then:

$$\Pr_{(x,y)\sim D}[\widetilde{y}_{new} \neq h^*(\boldsymbol{x}) \mid \widetilde{y} \text{ is rejected}] = \max((1),(3)) = C\max\left[O(\epsilon), C_1\right]^{\lambda}$$

For case 2, if we set $\delta = \max_{\boldsymbol{x}} \frac{\widetilde{\eta}_{\widetilde{y}}(\boldsymbol{x}|m_{\boldsymbol{x}},\widetilde{y})}{(\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}}-\tau_{\widetilde{y},m_{\boldsymbol{x}}})\eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})+\tau_{\widetilde{y},m_{\boldsymbol{x}}}}$, then by algebra $\frac{\widetilde{\eta}_{\widetilde{y}}(\boldsymbol{x}|m_{\boldsymbol{x}},\widetilde{y})/\delta-\tau_{\widetilde{y},m_{\boldsymbol{x}}}}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}}-\tau_{\widetilde{y},m_{\boldsymbol{x}}}} - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) < 0$, which implies that (2) is upper bounded by $C[O(\epsilon)]^{\lambda}$. Then:

$$\Pr_{(x,y)\sim D}[\widetilde{y}_{new} \neq h^*(\boldsymbol{x}) \mid \widetilde{y} \text{ is accepted}] = \max((2),(3)) = C\max\left[O(\epsilon), C_1\right]^{\lambda}$$

□

We give following several facts based on our theorem:

1. For symmetric noise $\tau_{ij} = \tau_{ji}, \forall i,j \in [N_c]$ and $f = \widetilde{\eta}$, we have $C_1 = 0$ since we always have $u_{\boldsymbol{x}} = m_{\boldsymbol{x}}$.
2. For binary case, the measure of case3 is 0. And $\frac{\tau_{\widetilde{y}_{\boldsymbol{x}},\widetilde{y}_{\boldsymbol{x}}}+\tau_{m_{\boldsymbol{x}},\widetilde{y}_{\boldsymbol{x}}}}{2-\tau_{\widetilde{y}_{\boldsymbol{x}},\widetilde{y}_{\boldsymbol{x}}}-\tau_{m_{\boldsymbol{x}},\widetilde{y}_{\boldsymbol{x}}}} = \frac{1+\tau_{m_{\boldsymbol{x}},\widetilde{y}_{\boldsymbol{x}}}-\tau_{\widetilde{y}_{\boldsymbol{x}},m_{\boldsymbol{x}}}}{1+\tau_{\widetilde{y}_{\boldsymbol{x}},m_{\boldsymbol{x}}}-\tau_{m_{\boldsymbol{x}},\widetilde{y}_{\boldsymbol{x}}}} = \frac{2-\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}}-\tau_{\widetilde{y}_{\boldsymbol{x}},m_{\boldsymbol{x}}}}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}}+\tau_{\widetilde{y}_{\boldsymbol{x}},m_{\boldsymbol{x}}}}$. If we take $\delta = \min\left(\frac{1+\tau_{10}-\tau_{01}}{1+\tau_{01}-\tau_{10}}, \frac{1+\tau_{01}-\tau_{10}}{1+\tau_{10}-\tau_{01}}\right)$. Then we have:

$$\Pr_{(\boldsymbol{x},y)\sim D}[\widetilde{y}_{new} \neq h^*(\boldsymbol{x})] \leq C[O(\epsilon)]^{\lambda}$$

3. From Lemma 1, we have $\Pr[\widetilde{y}_{new} = y] = \Pr[\widetilde{y}_{new} = h^*, h^* = y] + \Pr[\widetilde{y}_{new} = h^*, h^* \neq y] \geq \Pr[\widetilde{y}_{new} = h^*] + \Pr[h^* \neq y] \geq 1 - C\max\left[O(\epsilon), C_1\right]^{\lambda} - R^*$, where R^* is the Bayes risk.

Theorem 3. *Assume η and f satisfy the same conditions as Lemma 1. Also assume $\xi < \delta$. Let $\widetilde{y}_{new}$ be the output of the* `LRT-Correction` *with $(\boldsymbol{x},\widetilde{y})$, f, and the approximate $\hat{\delta}$. Assume that $\xi < \delta$, then for the constant $C_1 = \max_{\boldsymbol{x}} \frac{1-\tau_{m_{\boldsymbol{x}},u_{\boldsymbol{x}}}}{(\tau_{u_{\boldsymbol{x}},u_{\boldsymbol{x}}}-\tau_{m_{\boldsymbol{x}},u_{\boldsymbol{x}}})} - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})$:*

1. *Sensitivity Optimized Critical Value. Let $\delta = \min_{\boldsymbol{x}} \frac{(\tau_{\widetilde{y},\widetilde{y}}-\tau_{m_{\boldsymbol{x}},\widetilde{y}})\eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})+\tau_{m_{\boldsymbol{x}},\widetilde{y}}}{f_{m_{\boldsymbol{x}}}(\boldsymbol{x}|\widetilde{y},m_{\boldsymbol{x}})}$:*

$$\Pr_{(x,y)\sim D}[\widetilde{y}_{new} \neq h^*(\boldsymbol{x}) \mid \widetilde{y} \textit{ is rejected}] \leq C\max\left[O(\max(\epsilon,\xi)), C_1\right]^{\lambda}$$

2. *Specificity Optimized Critical Value. Let $\delta = \max_{\boldsymbol{x}} \frac{f_{\widetilde{y}}(\boldsymbol{x}|m_{\boldsymbol{x}},\widetilde{y})}{(\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}}-\tau_{\widetilde{y},m_{\boldsymbol{x}}})\eta_{s_{\boldsymbol{x}}}(\boldsymbol{x})+\tau_{\widetilde{y},m_{\boldsymbol{x}}}}$:*

$$\Pr_{(x,y)\sim D}[\widetilde{y}_{new} \neq h^*(\boldsymbol{x}) \mid \widetilde{y} \textit{ is accepted}] \leq C\max\left[O(\max(\epsilon,\xi)), C_1\right]^{\lambda}$$

Proof. The proof will be similar to the proof of Lemma 1, but we need to adjust the error introduced by picking $\hat{\delta}$. Recall that ξ and ϵ are both less than one.

If we pick $\hat{\delta}$ instead of δ, then for case 1 (where $\widetilde{y} = i$ is rejected) in Lemma 1, we have:

$$\begin{aligned}
&\Pr\left[\frac{f_i(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})}{f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})} < \hat{\delta}, \eta_i(\boldsymbol{x}) = \max \eta(\boldsymbol{x})\right] \leq \Pr\left[\frac{f_i(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})}{f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})} < \delta + \xi, \eta_i(\boldsymbol{x}) = \max \eta(\boldsymbol{x})\right] \\
&= \Pr\left[\eta_i(\boldsymbol{x}) > \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}), \eta_i(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) < \frac{\delta f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) - \tau_{m_{\boldsymbol{x}},i} + \xi f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})}{\tau_{ii} - \tau_{i,m_{\boldsymbol{x}}}} + \frac{\epsilon}{\tau_{ii} - \tau_{i,m_{\boldsymbol{x}}}}\right] \\
&\leq C\left[\frac{\delta f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}}) - \tau_{m_{\boldsymbol{x}},i}}{\tau_{ii} - \tau_{i,m_{\boldsymbol{x}}}} - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) + \frac{\xi f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})}{\tau_{ii} - \tau_{i,m_{\boldsymbol{x}}}} + \frac{\epsilon}{\tau_{ii} - \tau_{i,m_{\boldsymbol{x}}}}\right]^{\lambda} \\
&\leq C\left[\frac{\xi f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})}{\tau_{ii} - \tau_{i,m_{\boldsymbol{x}}}} + \frac{\epsilon + \xi}{\tau_{ii} - \tau_{m_{\boldsymbol{x}},i}}\right]^{\lambda} = C\left[O(max(\epsilon,\xi))\right]^{\lambda} \qquad (4)
\end{aligned}$$

We next analyze case 2 in Lemma 1 (where $\widetilde{y}_{\boldsymbol{x}}$ is accepted):

$$\Pr\left[\frac{f_i(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})}{f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})} \geq \hat{\delta}, \eta_{m_{\boldsymbol{x}}}(\boldsymbol{x}) = \max \eta(\boldsymbol{x})\right] \leq \Pr\left[\frac{f_i(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})}{f_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid i, m_{\boldsymbol{x}})} \geq \delta - \xi, \eta_{m_{\boldsymbol{x}}}(\boldsymbol{x}) = \max \eta(\boldsymbol{x})\right]$$

$$= \Pr\left[\eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) < \eta_{m_{\boldsymbol{x}}}(\boldsymbol{x}) < \eta_{m_{\boldsymbol{x}}}(\boldsymbol{x} \mid m_{\boldsymbol{x}}, i) \leq \frac{f_i(\boldsymbol{x} \mid m_{\boldsymbol{x}}, i)/(\delta - \xi) - \tau_{i,m_{\boldsymbol{x}}}}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}} - \tau_{i,m_{\boldsymbol{x}}}} + \frac{\epsilon}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}} - \tau_{i,m_{\boldsymbol{x}}}}\right] \tag{5}$$

$$= \Pr\left[0 < \eta_{m_{\boldsymbol{x}}}(\boldsymbol{x}) - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) < \frac{f_i(\boldsymbol{x} \mid m_{\boldsymbol{x}}, i)/\delta - \tau_{i,m_{\boldsymbol{x}}}}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}} - \tau_{i,m_{\boldsymbol{x}}}} - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) + \frac{\epsilon}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}} - \tau_{i,m_{\boldsymbol{x}}}} + \frac{\frac{\xi(f_i(\boldsymbol{x} \mid m_{\boldsymbol{x}}, i))}{\delta(\delta-\xi)}}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}} - \tau_{i,m_{\boldsymbol{x}}}}\right] \tag{6}$$

$$\leq C\left[\frac{f_i(\boldsymbol{x} \mid m_{\boldsymbol{x}}, i)/\delta - \tau_{i,m_{\boldsymbol{x}}}}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}} - \tau_{i,m_{\boldsymbol{x}}}} - \eta_{s_{\boldsymbol{x}}}(\boldsymbol{x}) + \frac{(1 + 1/\delta)\epsilon}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}} - \tau_{i,m_{\boldsymbol{x}}}} + \frac{\frac{\xi(f_i(\boldsymbol{x} \mid m_{\boldsymbol{x}}, i))}{\delta(\delta-\xi)}}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}} - \tau_{i,m_{\boldsymbol{x}}}}\right]^{\lambda} \tag{7}$$

$$\leq C\left[\frac{(1 + 1/\delta)\epsilon}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}} - \tau_{i,m_{\boldsymbol{x}}}} + \frac{\frac{\xi(f_i(\boldsymbol{x} \mid m_{\boldsymbol{x}}, i))}{\delta(\delta-\xi)}}{\tau_{m_{\boldsymbol{x}},m_{\boldsymbol{x}}} - \tau_{i,m_{\boldsymbol{x}}}}\right]^{\lambda} \tag{8}$$

Observe that $\frac{\xi}{\delta(\delta-\xi)} = \frac{\delta}{(\delta-\xi)}\frac{\xi}{\delta^2} = \left[1 + \xi + O(\xi^2)\right]\frac{\xi}{(1+\delta)^2}$. The final equality uses Taylor expansion if $\xi < \delta$.

Case 3 in Lemma 1 will not be affected by the choice of $\hat{\delta}$, which completes the proof. $\square$