A Omitted Proofs

For some of our proofs we will need the following bound:

Theorem A.1 (Hoeffding's inequality (Hoeffding (1963))). *Let $Z_1, \ldots, Z_n$ be independent random variables with $Z_i \in [a, b]$, for all $i \in [n]$. Then, for all $\epsilon > 0$,*

$$\mathbf{Pr}\left[\left|\frac{1}{n}\sum_{i=1}^{n}(Z_i - \mathbb{E}[Z_i])\right| \geq \epsilon\right] \leq 2\exp\left(-\frac{2n\epsilon^2}{(b-a)^2}\right). \tag{7}$$

A.1 Proof of Lemma 3.2

Proof. Consider some $t \in [0, t_0]$, and let $\mathcal{D}_{\mathbf{x}}$ denote the marginal distribution on the unlabeled points. By definition of the Tsybakov noise condition, the instance space $\mathcal{X}$ may be partitioned into regions $\mathcal{X}_{good}$ and $\mathcal{X}_{bad}$ such that

- $\mathbf{Pr}_{\mathbf{x}\sim\mathcal{D}_{\mathbf{x}}}[\mathbf{x} \in \mathcal{X}_{good}] \geq 1 - At^{\frac{\alpha}{1-\alpha}}$, and $\eta(\mathbf{x}) \leq \frac{1}{2} - t$ almost surely for all $\mathbf{x} \in \mathcal{X}_{good}$. The points in $\mathcal{X}_{good}$ should be thought of as being corrupted with Massart noise;
- $\mathbf{Pr}_{\mathbf{x}\sim\mathcal{D}_{\mathbf{x}}}[\mathbf{x} \in \mathcal{X}_{bad}] \leq At^{\frac{\alpha}{1-\alpha}}$. The points in $\mathcal{X}_{bad}$ may have flipping probabilities arbitrarily close to 1/2.

As a result, it follows that

$$\int_{\mathcal{X}}(1-2\eta(\mathbf{x}))\mathcal{D}_{\mathbf{x}}(\mathbf{x})d\mathbf{x} = \int_{\mathcal{X}_{good}}(1-2\eta(\mathbf{x}))\mathcal{D}_{\mathbf{x}}(\mathbf{x})d\mathbf{x} + \overbrace{\int_{\mathcal{X}_{bad}}(1-2\eta(\mathbf{x}))\mathcal{D}_{\mathbf{x}}(\mathbf{x})d\mathbf{x}}^{>0} \tag{8}$$

$$> 2t\int_{\mathcal{X}_{good}}\mathcal{D}_{\mathbf{x}}(\mathbf{x})d\mathbf{x} \tag{9}$$

$$\geq 2t(1 - At^{\frac{\alpha}{1-\alpha}}), \tag{10}$$

where in the first line we used that $\eta(\mathbf{x}) < \frac{1}{2}$ for all $\mathbf{x} \in \mathcal{X}$. As a result, we obtain that

$$\mathcal{M}^{-1} \geq \sup_{t\in[0,t_0]}\left\{2t(1 - At^{\frac{\alpha}{1-\alpha}})\right\}. \tag{11}$$

Finally, it is easy to verify that

$$\sup_{t\in[0,t_0]}\left\{2t(1 - At^{\frac{\alpha}{1-\alpha}})\right\} = \begin{cases} 2\alpha\left(\frac{1-\alpha}{A}\right)^{\frac{1-\alpha}{\alpha}} & \text{if } t^* \leq t_0, \\ 2t_0\left(1 - At_0^{\frac{\alpha}{1-\alpha}}\right) & \text{if } t^* > t_0. \end{cases}$$

We should mention that when $t^* > t_0$, it follows that $At_0^{\frac{\alpha}{1-\alpha}} \neq 1$. ◻

A.2 Proof of Lemma 4.1

Proof. First of all, we have that

$$\mathbf{Pr}_{\mathbf{x}\sim\mathcal{D}'_{\mathbf{x}}}[h(\mathbf{x}) \neq f(\mathbf{x})] = \frac{1}{Z}\mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}_{\mathbf{x}}}[(1-2\eta(\mathbf{x}))\mathbb{1}\{h(\mathbf{x}) \neq f(\mathbf{x})\}] \tag{12}$$

$$\geq \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}_{\mathbf{x}}}[(1-2\eta(\mathbf{x}))\mathbb{1}\{h(\mathbf{x}) \neq f(\mathbf{x})\}], \tag{13}$$

where the last inequality follows from $Z \leq 1$. Moreover, we obtain that

$$\mathbf{Pr}_{(\mathbf{x},y)\sim\mathcal{D}}[h(\mathbf{x}) \neq y] = \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}_{\mathbf{x}}}[(1-\eta(\mathbf{x}))\mathbb{1}\{h(\mathbf{x}) \neq f(\mathbf{x})\}] + \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}_{\mathbf{x}}}[\eta(\mathbf{x})\mathbb{1}\{h(\mathbf{x}) = f(\mathbf{x})\}] \tag{14}$$

$$= \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}_{\mathbf{x}}}[\eta(\mathbf{x})] + \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}_{\mathbf{x}}}[(1-2\eta(\mathbf{x}))\mathbb{1}\{h(\mathbf{x}) \neq f(\mathbf{x})\}] \tag{15}$$

$$\leq \mathrm{OPT} + \epsilon, \tag{16}$$

where we used that $\mathrm{OPT} = \mathbb{E}_{\mathbf{x}\sim\mathcal{D}_{\mathbf{x}}}[\eta(\mathbf{x})]$ and $\mathbf{Pr}_{\mathbf{x}\sim\mathcal{D}'_{\mathbf{x}}}[h(\mathbf{x} \neq f(\mathbf{x}))] \leq \epsilon$. □

A.3 Proof of Lemma 4.2

Proof. It follows that

$$\mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}}[\psi(\mathbf{x},y)] = \int_{\mathcal{X}} \phi(\mathbf{x})f(\mathbf{x})(1-2\eta(\mathbf{x}))\mathcal{D}_{\mathbf{x}}(\mathbf{x})d\mathbf{x} \tag{17}$$

$$= Z \int_{\mathcal{X}} \phi(\mathbf{x})f(\mathbf{x})\mathcal{D}'_{\mathbf{x}}(\mathbf{x})d\mathbf{x} \tag{18}$$

$$= Z \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}'_{\mathbf{x}}}[\phi(\mathbf{x})f(\mathbf{x})] \tag{19}$$

$$= Z \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}'}[\psi(\mathbf{x},y)]. \tag{20}$$

□

A.4 Proof of Theorem 4.3

Proof. First of all, given that $\mathcal{A}$ efficiently learns up to an ϵ error the concept class $\mathcal{C}$, it follows that $q = \mathrm{poly}(d, 1/\epsilon)$ and $1/\tau = \mathrm{poly}(d, 1/\epsilon)$. For some iteration in the main loop of the algorithm, $\tilde{Z}$ will be such that $Z \leq \tilde{Z} \leq Z + \tau'$. For this particular $\tilde{Z}$, it follows that $|1/Z - 1/\tilde{Z}| \leq \tau'/Z^2 \leq \tau' C^2 = \tau/2$, where we used that $Z \geq 1/C$.

Now consider any correlational statistical query $\psi(\mathbf{x}, y)$; we have to establish that when our guess for parameter Z is close to the actual value, every query of algorithm $\mathcal{A}$ is simulated correctly with high probability. Indeed, Lemma 4.2 implies that

$$\left| \frac{1}{\tilde{Z}} \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}}[\psi(\mathbf{x},y)] - \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}'_{\mathbf{x}}}[\psi(\mathbf{x}, f(\mathbf{x}))] \right| = \left| \frac{1}{\tilde{Z}} - \frac{1}{Z} \right| \left| \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}}[\psi(\mathbf{x},y)] \right| \leq \tau/2, \tag{21}$$

where $\mathcal{D}'$ is defined as in Lemma 4.2. Moreover, let $\widehat{\mathbb{E}}_{\mathcal{D}}[\psi(\mathbf{x},y)]$ be the empirical estimate of $\mathbb{E}_{(\mathbf{x},y)\sim\mathcal{D}}[\psi(\mathbf{x},y)]$ formed from $\mathcal{O}\left(C^2 \log(q/\delta)/\tau^2\right)$ samples. Given that $|\psi(\mathbf{x},y)| \leq 1$, Hoeffding's inequality implies that with probability at least $1 - \delta/q$,

$$\left| \mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}}[\psi(\mathbf{x},y)] - \widehat{\mathbb{E}}_{\mathcal{D}}[\psi(\mathbf{x},y)] \right| \leq \frac{\tau}{2C} \leq \frac{\tau\tilde{Z}}{2}. \tag{22}$$

As a result, combining (21) and (22) yields that with probability at least $1 - \delta/q$,

$$\left| \frac{1}{\tilde{Z}} \widehat{\mathbb{E}}_{\mathcal{D}}[\psi(\mathbf{x},y)] - \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}'_{\mathbf{x}}}[\psi(\mathbf{x}, f(\mathbf{x}))] \right| \leq \frac{\tau'}{Z^2} \leq \tau. \tag{23}$$

By the union bound, we obtain that for the $\tilde{Z}$ that satisfies $Z \leq \tilde{Z} \leq Z + \tau'$, all of the q CSQ queries made by algorithm $\mathcal{A}$ are answered correctly up to error τ with probability at least $1-\delta$. Then, for this particular iteration the output hypothesis h of algorithm $\mathcal{A}$ satisfies $\mathbf{Pr}_{\mathbf{x}\sim\mathcal{D}'_{\mathbf{x}}}[h(\mathbf{x}) \neq f(\mathbf{x})] \leq \epsilon$, which – by Lemma 4.1 – implies that $\mathbf{Pr}_{(\mathbf{x},y)\sim\mathcal{D}}[h(\mathbf{x}) \neq y] \leq \mathrm{OPT} + \epsilon$. Finally, let $\widehat{\mathbf{Pr}}_{\mathcal{D}}[h(\mathbf{x}) \neq y]$ be the empirical estimate of $\mathbf{Pr}_{(\mathbf{x},y)\sim\mathcal{D}}[h(\mathbf{x}) \neq y]$. If we invoke $\mathcal{O}\left(\log(1/\delta)/\epsilon^2\right)$ samples, we obtain that with probability at least $1 - \delta$,

$$\left|\widehat{\mathbf{Pr}}_{\mathcal{D}}[h(\mathbf{x}) \neq y] - \mathbf{Pr}_{(\mathbf{x},y)\sim\mathcal{D}}[h(\mathbf{x}) \neq y]\right| \leq \epsilon. \tag{24}$$

Thus, by the union bound $\mathcal{O}\left(\log(N/\delta)/\epsilon^2\right)$ samples suffice to guarantee that the estimation error is up to ϵ in every iteration with probability at least $1-\delta$, where $N = \mathcal{O}(C^2/\tau)$ is the number of iterations of the main loop in the algorithm. Consequently, the output of the algorithm h satisfies, with probability at least $1-2\delta$, $\mathbf{Pr}_{(\mathbf{x},y)\sim\mathcal{D}}[h(\mathbf{x}) \neq y] \leq \mathrm{OPT} + 3\epsilon$. Finally, rescaling ϵ and δ concludes the proof. □

A.5 Proof of Lemma 5.2

Proof. If f represents the target function, the claim follows from the following observation:

$$\mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}'_{\mathbf{x}}}[\psi(\mathbf{x}, f(\mathbf{x}))] = \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}'_{\mathbf{x}}}\left[\psi(\mathbf{x},-1)\frac{1-f(\mathbf{x})}{2} + \psi(\mathbf{x},1)\frac{1+f(\mathbf{x})}{2}\right] \tag{25}$$

$$= \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}'_{\mathbf{x}}}\left[\frac{\psi(\mathbf{x},1)-\psi(\mathbf{x},-1)}{2}f(\mathbf{x})\right] + \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}'_{\mathbf{x}}}\left[\frac{\psi(\mathbf{x},1)+\psi(\mathbf{x},-1)}{2}\right]. \tag{26}$$

□

A.6 Proof of Lemma 5.3

Proof. Let (ϕ', τ) represent the target independent statistical query. In the interest of simplifying our argument we notice that

$$\mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}'_{\mathbf{x}}}[\phi'(\mathbf{x})] = \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}'_{\mathbf{x}}}\left[-1 + 2\frac{1+\phi'(\mathbf{x})}{2}\right] = -1 + 2\mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}'_{\mathbf{x}}}[\phi(\mathbf{x})], \tag{27}$$

where $\phi(\mathbf{x}) = (1+\phi'(\mathbf{x}))/2$. Thus, it suffices to simulate the statistical query $(\phi, \tau/2)$ on $\mathcal{D}'_{\mathbf{x}}$, where ϕ takes values in $[0,1]$. If $Z = \mathcal{M}^{-1} = \mathbb{E}_{\mathcal{D}_{\mathbf{x}}}[1-2\eta(\mathbf{x})]$, we have that

$$\mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}'_{\mathbf{x}}}[\phi(\mathbf{x})] = \frac{1}{Z}\int_{\mathcal{X}} \phi(\mathbf{x})(1-2\eta(\mathbf{x}))\mathcal{D}_{\mathbf{x}}(\mathbf{x})d\mathbf{x} = \frac{1}{Z}\mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}_{\mathbf{x}}}[\phi(\mathbf{x})(1-2\eta(\mathbf{x}))]. \tag{28}$$

Let $\widehat{Z}$ be the empirical estimate of $\mathbb{E}_{\mathcal{D}_{\mathbf{x}}}[1-2\eta(\mathbf{x})]$ formed from $\mathcal{O}\left(\log(1/\delta)/(\tau')^2\right)$ samples of $\mathrm{EX}^{\eta}(f, \mathcal{D}_{\mathbf{x}}, \eta)$, for some $\delta > 0$ and $\tau' := \tau/(2C)$. Given that $0 \leq 1-2\eta(\mathbf{x}) \leq 1, \forall \mathbf{x} \in \mathcal{X}$, Hoeffding's inequality implies that $|\widehat{Z} - Z| < \tau'/2$, with probability at least $1-\delta$. Thus, if we let $\widehat{Z} := \widehat{Z} + \tau'/2$, we obtain that $Z < \widehat{Z} < Z + \tau'$, with probability at least $1-\delta$. Furthermore, let $\widehat{\mathbb{E}}_{\mathcal{D}_{\mathbf{x}}}[\phi(\mathbf{x})(1-2\eta(\mathbf{x}))]$ be the empirical estimate of $\mathbb{E}_{\mathcal{D}_{\mathbf{x}}}[\phi(\mathbf{x})(1-2\eta(\mathbf{x}))]$ formed from $\mathcal{O}(\log(1/\delta)/(\tau')^2)$ of $\mathrm{EX}^{\eta}(f, \mathcal{D}_{\mathbf{x}}, \eta)$. If we increment the estimate by $\tau'/2$ we can again guarantee that $\mathbb{E}_{\mathcal{D}_{\mathbf{x}}}[\phi(\mathbf{x})(1-2\eta(\mathbf{x}))] < \widehat{\mathbb{E}}_{\mathcal{D}_{\mathbf{x}}}[\phi(\mathbf{x})(1-2\eta(\mathbf{x}))] < \mathbb{E}_{\mathcal{D}_{\mathbf{x}}}[\phi(\mathbf{x})(1-2\eta(\mathbf{x}))] + \tau'$, with probability at least $1-\delta$. Indeed, given that $0 \leq \phi(\mathbf{x})(1-2\eta(\mathbf{x})) \leq 1, \forall \mathbf{x} \in \mathcal{X}$, we can directly apply Hoeffding's inequality. As a result, with probability at least $1-2\delta$ we have that

$$\frac{\widehat{\mathbb{E}}_{\mathcal{D}_{\mathbf{x}}}[\phi(\mathbf{x})(1-2\eta(\mathbf{x}))]}{\widehat{Z}} < \frac{\mathbb{E}_{\mathcal{D}_{\mathbf{x}}}[\phi(\mathbf{x})(1-2\eta(\mathbf{x}))] + \tau'}{Z} \leq \mathop{\mathbb{E}}_{\mathcal{D}'_{\mathbf{x}}}[\phi(\mathbf{x})] + \tau' C = \mathop{\mathbb{E}}_{\mathcal{D}'_{\mathbf{x}}}[\phi(\mathbf{x})] + \frac{\tau}{2}, \tag{29}$$

$$\frac{\widehat{\mathbb{E}}_{\mathcal{D}_{\mathbf{x}}}[\phi(\mathbf{x})(1-2\eta(\mathbf{x}))]}{\widehat{Z}} > \frac{\mathbb{E}_{\mathcal{D}_{\mathbf{x}}}[\phi(\mathbf{x})(1-2\eta(\mathbf{x}))]}{Z+\tau'} \geq \frac{1}{1+\tau/2}\mathop{\mathbb{E}}_{\mathcal{D}'_{\mathbf{x}}}[\phi(\mathbf{x})], \tag{30}$$

where in the final bound we used that $\tau' \leq \tau Z/2$. Thus, it follows from (30) that

$$\frac{\widehat{\mathbb{E}}_{\mathcal{D}_{\mathbf{x}}}[\phi(\mathbf{x})(1-2\eta(\mathbf{x}))]}{\widehat{Z}} - \mathop{\mathbb{E}}_{\mathcal{D}'_{\mathbf{x}}}[\phi(\mathbf{x})] > \mathop{\mathbb{E}}_{\mathcal{D}'_{\mathbf{x}}}[\phi(\mathbf{x})]\left(\frac{1}{1+\tau/2} - 1\right) \geq -\mathop{\mathbb{E}}_{\mathcal{D}'_{\mathbf{x}}}[\phi(\mathbf{x})]\frac{\tau}{2} \geq -\frac{\tau}{2}, \tag{31}$$

since $\tau > 0$ and $0 \leq \mathbb{E}_{\mathcal{D}'_{\mathbf{x}}}[\phi(\mathbf{x})] \leq 1$. As a result, if we combine (29) and (31) we obtain that

$$-\frac{\tau}{2} < \frac{\widehat{\mathbb{E}}_{\mathcal{D}_{\mathbf{x}}}[\phi(\mathbf{x})(1-2\eta(\mathbf{x}))]}{\widehat{Z}} - \mathop{\mathbb{E}}_{\mathcal{D}'_{\mathbf{x}}}[\phi(\mathbf{x})] < \frac{\tau}{2}, \tag{32}$$

with probability at least $1-2\delta$; finally, rescaling $\delta := \delta/2$ concludes the proof. □

A.7 Proof of Lemma 5.4

Proof. Let $Z = \mathcal{M}^{-1}$ and $\psi(\mathbf{x}, f(\mathbf{x})) = \phi(\mathbf{x})f(\mathbf{x})$ the input query. Every correlational statistical query on distribution $\mathcal{D}'_{\mathbf{x}}$ can be expressed as

$$\mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}'_{\mathbf{x}}}[\phi(\mathbf{x})f(\mathbf{x})] = \frac{1}{Z}\int_{\mathcal{X}} \phi(\mathbf{x})f(\mathbf{x})(1-2\eta(\mathbf{x}))\mathcal{D}_{\mathbf{x}}(\mathbf{x})d\mathbf{x} = \frac{1}{Z}\mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}}[\phi(\mathbf{x})y]. \tag{33}$$

Let $\widehat{Z}$ be the empirical estimate of Z from $\mathcal{O}(\log(1/\delta)/(\tau')^2)$ samples of $\mathrm{EX}^{\eta}(f, \mathcal{D}_{\mathbf{x}}, \eta)$, for some $\delta > 0$ and $\tau' := \tau/(2C^2)$. If we increment our estimate by $\tau'/2$, it follows that $Z < \widehat{Z} < Z + \tau'$ with probability at least $1-\delta$. Thus, we obtain that

$$\left| \frac{1}{Z}\mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}}[\phi(\mathbf{x})y] - \frac{1}{\widehat{Z}}\mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}}[\phi(\mathbf{x})y] \right| \le \frac{\tau'}{Z^2} \le \tau' C^2 = \frac{\tau}{2}. \tag{34}$$

Moreover, let $\widehat{E}_{\mathcal{D}}[\phi(\mathbf{x})y]$ the empirical estimate of $\mathbb{E}_{\mathcal{D}}[\phi(\mathbf{x})y]$. For $Z < \widehat{Z}$, Hoeffding's inequality implies that $\mathcal{O}(C^2 \log(1/\delta)/\tau^2)$ samples suffice so that

$$\left| \frac{1}{\widehat{Z}}\widehat{\mathbb{E}}_{\mathcal{D}}[\phi(\mathbf{x})y] - \frac{1}{\widehat{Z}}\mathop{\mathbb{E}}_{(\mathbf{x},y)\sim\mathcal{D}}[\phi(\mathbf{x})y] \right| < \frac{\tau}{2\widehat{Z}C} < \frac{\tau}{2ZC} < \frac{\tau}{2}, \tag{35}$$

with probability at least $1-\delta$. Thus, combining (34) and (35) we obtain that with probability at least $1-2\delta$,

$$\left| \frac{1}{\widehat{Z}}\widehat{E}_{\mathcal{D}}[\phi(\mathbf{x})y] - \mathop{\mathbb{E}}_{\mathcal{D}'_{\mathbf{x}}}[\phi(\mathbf{x})f(\mathbf{x})] \right| < \tau. \tag{36}$$

□

B Optimality in the Realizable Instance

In this section we analyze whether obtaing a hypothesis h such that $\mathrm{err}_{\mathcal{D}}(h) \le \mathrm{OPT} + \epsilon$ implies that $\mathbf{Pr}_{\mathcal{D}_{\mathbf{x}}}[h(\mathbf{x}) \neq f(\mathbf{x})] \le \epsilon'$, for some ϵ' that depends polynomially on ϵ. To be more precise, we show that this is indeed the case in the Massart as well as the Tsybakov model, but as we will see it does not hold in general.

Massart Model. Consider a hypothesis h such that $\mathrm{err}_{\mathcal{D}}(h) \le \mathrm{OPT} + \epsilon$, for any $\epsilon > 0$. Then, given that $\eta(\mathbf{x}) \le \gamma$, it follows that

$$\begin{aligned} \mathrm{err}_{\mathcal{D}}(h) &= \mathrm{OPT} + \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}_{\mathbf{x}}}[(1-2\eta(\mathbf{x}))\mathbb{1}\{h(\mathbf{x}) \neq f(\mathbf{x})\}] &(37)\\ &\ge \mathrm{OPT} + (1-2\gamma)\,\mathbf{Pr}_{\mathbf{x}\sim\mathcal{D}_{\mathbf{x}}}[f(\mathbf{x}) \neq h(\mathbf{x})]. &(38) \end{aligned}$$

Thus, we obtain that

$$\mathbf{Pr}_{\mathbf{x}\sim\mathcal{D}_{\mathbf{x}}}[f(\mathbf{x}) \neq h(\mathbf{x})] \le \frac{\epsilon}{1-2\gamma}. \tag{39}$$

As a result, it suffices to select $\epsilon = \epsilon'(1-2\gamma)$ to guarantee ϵ' excess error in the underlying realizable instance.

Tsybakov Model. Again, consider a hypothesis h such that $\mathrm{err}_{\mathcal{D}}(h) \leq \mathrm{OPT} + \epsilon$, for any $\epsilon > 0$, and fix some $t \in [0, t_0]$. Employing similar ideas to Lemma 3.2 yields that

$$\mathrm{err}_{\mathcal{D}}(h) = \mathrm{OPT} + \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{D}_{\mathbf{x}}}[(1-2\eta(\mathbf{x}))\mathbb{1}\{h(\mathbf{x}) \neq f(\mathbf{x})\}] \tag{40}$$

$$\geq \mathrm{OPT} + \int_{\mathcal{X}_{good}} (1-2\eta(\mathbf{x}))\mathbb{1}\{h(\mathbf{x}) \neq f(\mathbf{x})\}\mathcal{D}_{\mathbf{x}}(\mathbf{x})d\mathbf{x} \tag{41}$$

$$\geq \mathrm{OPT} + 2t \int_{\mathcal{X}_{good}} \mathbb{1}\{h(\mathbf{x}) \neq f(\mathbf{x})\}\mathcal{D}_{\mathbf{x}}(\mathbf{x})d\mathbf{x}, \tag{42}$$

where $\mathcal{X}_{good}$ is defined as in Lemma 3.2. Moreover, given that $\mathbf{Pr}_{\mathcal{D}_{\mathbf{x}}}[\mathbf{x} \in \mathcal{X}_{good}] \geq 1 - At^{\frac{\alpha}{1-\alpha}}$, we obtain that

$$\mathbf{Pr}_{\mathbf{x}\sim\mathcal{D}_{\mathbf{x}}}[h(\mathbf{x}) \neq f(\mathbf{x})] \leq \frac{\epsilon}{2t} + At^{\frac{\alpha}{1-\alpha}}. \tag{43}$$

Therefore, in order to get $\mathbf{Pr}_{\mathbf{x}\sim\mathcal{D}_{\mathbf{x}}}[h(\mathbf{x}) \neq f(\mathbf{x})] \leq \epsilon'$, for any $\epsilon' > 0$, it suffices to select ϵ such that

$$\epsilon = \sup_{t\in[0,t_0]} \left\{2t\epsilon' - 2At^{\frac{1}{1-\alpha}}\right\}. \tag{44}$$

In particular, it follows that

$$\sup_{t\in[0,t_0]} \left\{2t\epsilon' - 2At^{\frac{1}{1-\alpha}}\right\} = \begin{cases} 2(\epsilon')^{\frac{1}{\alpha}} \left(\frac{1-\alpha}{A}\right)^{\frac{1-\alpha}{\alpha}} - 2A\left(\epsilon'\frac{1-\alpha}{A}\right)^{\frac{1}{\alpha}} & \text{if } t^* \leq t_0, \\ 2t_0\epsilon' - 2At_0^{\frac{1}{1-\alpha}} & \text{if } t^* > t_0, \end{cases}$$

where

$$t^* = \left(\epsilon'\frac{1-\alpha}{A}\right)^{\frac{1-\alpha}{\alpha}}. \tag{45}$$

On the other hand, consider the following noise function:

Definition B.1. *A noise function $\eta(\mathbf{x})$ satisfies a β-clean condition if there exists a region $\mathcal{X}_{clean} \subseteq \mathcal{X}$ such that*

- $\mathbf{Pr}_{\mathbf{x}\sim\mathcal{D}_{\mathbf{x}}}[\mathbf{x} \in \mathcal{X}_{clean}] \geq \beta$;
- $\eta(\mathbf{x}) = 0, \forall \mathbf{x} \in \mathcal{X}_{clean}$.

This noise condition allows a $1-\beta$ fraction of the probability mass to be corrupted with noise arbitrarily close to $1/2$.

Lemma B.2. *The magnitude of a β-clean noise with respect to any distribution $\mathcal{D}_{\mathbf{x}}$ is upper-bounded by $1/\beta$.*

Proof. It follows that

$$\mathcal{M}^{-1} = \int_{\mathcal{X}} (1-2\eta(\mathbf{x}))\mathcal{D}_{\mathbf{x}}(\mathbf{x})d\mathbf{x} \geq \int_{\mathcal{X}_{clean}} \mathcal{D}_{\mathbf{x}}(\mathbf{x})d\mathbf{x} \geq \beta. \tag{46}$$

□

However, in this particular noise model a guarantee in the noisy distribution does not necessarily translate in the realizable instance. Indeed, assume that $\mathcal{D}_{\mathbf{x}}$ is the uniform distribution on $\mathbb{B}_2 = \{\mathbf{x} \in \mathbb{R}^2 : ||\mathbf{x}||_2 \leq 1\}$. We consider a partition of $\mathbb{B}_2$ into $\mathcal{X}^r_{clean}$, $\mathcal{X}^\ell_{clean}$, and the region $\mathbb{B}_2 \setminus (\mathcal{X}^r_{clean} \cup \mathcal{X}^\ell_{clean})$, as indicated in Figure 1, and

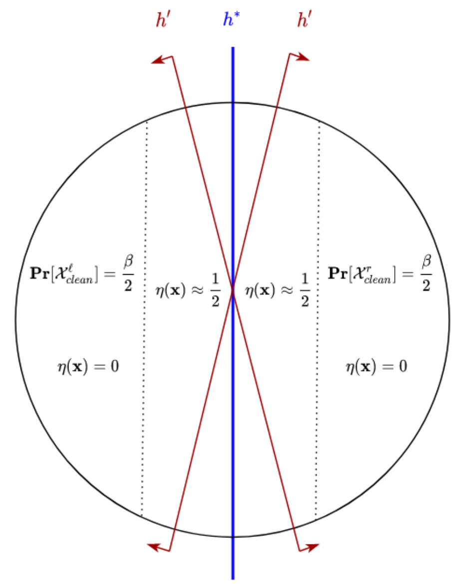

Figure 1: The geometry of our example; here h^* represents the optimal classifier.

we let $\mathbf{Pr}_{\mathcal{D}_\mathbf{x}}[\mathbf{x} \in \mathcal{X}^{\ell}_{clean}] = \mathbf{Pr}_{\mathcal{D}_\mathbf{x}}[\mathbf{x} \in \mathcal{X}^{r}_{clean}] = \frac{\beta}{2}$. In addition, we let $\eta(\mathbf{x}) = 0, \forall \mathbf{x} \in \mathcal{X}^{r}_{clean} \cup \mathcal{X}^{\ell}_{clean}$, while for the rest of the probability mass we let $\eta(\mathbf{x}) = \frac{1}{2} - \rho$, for some $\rho > 0$.

The problem that arises is that in the limit of $\rho \to 0$, $\mathrm{err}_{\mathcal{D}}(h') \to \mathrm{err}_{\mathcal{D}}(h^*) = \mathrm{OPT}$, for any h' as in Figure 1. Yet, it is clear that in the realizable instance the error of h' can be very far from the optimal. Nonetheless, it should be noted that a hypothesis h such that $\mathrm{err}_{\mathcal{D}}(h) \leq \mathrm{OPT} + \epsilon$ would classify correctly the clean data even in the presence of intense noise, a result that appears to be non-trivial and of independent interest.