A. Proof of Theorem 5.1

The proof of Theorem 5.1 is as follows.

Proof. Define the input set of disagreement with margin θ:

$$\mathcal{A}_\theta(\eta, \eta_{\text{corr}}) := \left(\left\{ \eta(x) \leq \frac{1}{2} - \theta \right\} \cap \left\{ \eta_{\text{corr}}(x) \geq \frac{1}{2} + \theta \right\} \right) \tag{13}$$

$$\cup \left(\left\{ \eta(x) \geq \frac{1}{2} + \theta \right\} \cap \left\{ \eta_{\text{corr}}(x) \leq \frac{1}{2} - \theta \right\} \right). \tag{14}$$

We can write $\mathcal{A}_0(\eta, \eta_{\text{corr}})$ as a union of such sets: $\mathcal{A}_0(\eta) = \bigcup_{\theta > 0} \mathcal{A}_\theta(\eta)$, and hence

$$\lim_{\theta \to 0} \left\{ \mu \left(\mathcal{A}_\theta(\eta, \eta_{\text{corr}}) \right) \right\} = \mu \left(\mathcal{A}_0(\eta, \eta_{\text{corr}}) \right) > 0.$$

Now, take some $\theta > 0$ s.t. $\mathcal{A}_0(\eta, \eta_{\text{corr}}) > 0$. Lemma A.1 below will show that

$$\mathcal{E}\left(\phi; \mathbb{P}_{(\mu,\eta)}\right) + \mathcal{E}\left(\phi; \mathbb{P}_{(\mu,\eta_{\text{corr}})}\right) \geq \theta \cdot \mu \left(\mathcal{A}_\theta(\eta, \eta_{\text{corr}}) \right) > 0. \tag{15}$$

Since $\hat{\phi}$ is consistent with $\mathbb{P}_{train} = \mathbb{P}_{test} = \mathbb{P}_{(\mu,\eta_{\text{corr}})}$, we have $\lim_{n\to\infty} \mathcal{E}\left(\hat{\phi}_n; \mathbb{P}_{(\mu,\eta_{\text{corr}})}\right) = 0$. Hence, from eq. (15) it follows that $\limsup_{n\to\infty} \mathcal{E}\left(\hat{\phi}_n; \mathbb{P}_{(\mu,\eta)}\right) \geq \theta \cdot \mu \left(\mathcal{A}_\theta(\eta, \eta_{\text{corr}}) \right) > 0$. That is, $\hat{\phi}$ is inconsistent when trained with train distribution $\mathbb{P}_{(\mu,\eta_{\text{corr}})}$ and tested on distribution $\mathbb{P}_{(\mu,\eta)}$. □

It remains to prove the lemma used in the proof above.

Lemma A.1. *Let μ be a Borel probability measure on $\mathcal{X}$. Given $\eta : \mathcal{X} \to [0,1]$ and $\theta > 0$ consider the set $\mathcal{A}_\theta(\eta, \eta_{corr}) \subseteq \mathcal{X}$ as defined in eq. (14). Then given any classifier $\phi : \mathcal{X} \to \{0,1\}$ we have $\mathcal{E}\left(\phi; \mathbb{P}_{(\mu,\eta)}\right) + \mathcal{E}\left(\phi; \mathbb{P}_{(\mu,\eta_{corr})}\right) \geq \theta \cdot \mu \left(\mathcal{A}_\theta(\eta, \eta_{corr}) \right)$.*

Proof. Recall that, for any regression function $\tilde{\eta} : \mathcal{X} \to [0,1]$ the excess risk can be written as: $\mathcal{E}\left(\phi; \mathbb{P}_{(\mu,\tilde{\eta})}\right) =$

$$\int \left| \tilde{\eta}(x) - \frac{1}{2} \right| \cdot \mathbb{1}\left\{ \left(\tilde{\eta}(x) - \frac{1}{2} \right) \left(\phi(x) - \frac{1}{2} \right) < 0 \right\} d\mu(x). \tag{16}$$

Now if $x \in \mathcal{A}_\theta(\eta, \eta_{\text{corr}})$ then $\left(\eta(x) - \frac{1}{2}\right)\left(\eta_{\text{corr}}(x) - \frac{1}{2}\right) < 0$ so for both possible values $\phi(x) \in \{0,1\}$ we have

$$\mathbb{1}\left\{ \left(\eta(x) - \frac{1}{2} \right) \left(\phi(x) - \frac{1}{2} \right) < 0 \right\} + \mathbb{1}\left\{ \left(\eta_{\text{corr}}(x) - \frac{1}{2} \right) \left(\phi(x) - \frac{1}{2} \right) < 0 \right\} = 1.$$

Moreover, if $x \in \mathcal{A}_\theta(\eta, \eta_{\text{corr}})$ then $\min \left\{ \left|\eta(x) - \frac{1}{2}\right|, \left|\eta_{\text{corr}}(x) - \frac{1}{2}\right| \right\} \geq \theta$ and so

$$\left|\eta(x) - \frac{1}{2}\right| \cdot \mathbb{1}\left\{ \left(\eta(x) - \frac{1}{2} \right) \left(\phi(x) - \frac{1}{2} \right) < 0 \right\} + \left|\eta_{\text{corr}}(x) - \frac{1}{2}\right| \cdot \mathbb{1}\left\{ \left(\eta_{\text{corr}}(x) - \frac{1}{2} \right) \left(\phi(x) - \frac{1}{2} \right) < 0 \right\} \geq \theta. \tag{17}$$

Integrating with respect to μ and applying (16) to both $\mathbb{P}_{(\mu,\eta)}$ and $\mathbb{P}_{(\mu,\eta_{\text{corr}})}$ gives the conclusion of the lemma. □

B. Proof of Lemma 3.1

Proof. Given $\hat{a}, a \in [-1, 1]$, $b, \hat{b} > 0$ with $|\hat{b} - b| \leq b/2$, and $a/b \in [0, 1]$,

$$\left|\frac{\hat{a}}{\hat{b}} - \frac{a}{b}\right| = \frac{1}{\hat{b}} \cdot \left|(\hat{a} - a) + \frac{a}{b} \cdot (b - \hat{b})\right| \leq \frac{2}{b}\left(|\hat{a} - a| + \left|\frac{a}{b}\right| \cdot |\hat{b} - b|\right) \leq \frac{4}{b} \cdot \max\left\{|\hat{a} - a|, |\hat{b} - b|\right\}, \tag{18}$$

where we have used the fact that $\hat{b} \geq b/2$. By the definition of $\hat{\eta}(x)$ together with eq. (1) we have

$$\hat{\eta}(x) := \frac{\hat{\eta}_{\text{corr}}(x) - \hat{p}_0}{1 - \hat{p}_0 - \hat{p}_1} \qquad \text{and} \qquad \eta(x) = \frac{\eta_{\text{corr}}(x) - p_0}{1 - p_0 - p_1}.$$

Now take $\hat{a} = \hat{\eta}_{\text{corr}}(x) - \hat{p}_0$, $a = \eta_{\text{corr}}(x) - p_0$, $\hat{b} = 1 - \hat{p}_0 - \hat{p}_1$ and $b = 1 - p_0 - p_1$. Given the assumptions that $p_0 + p_1 < 1$, so $b > 0$ and $\max\{|\hat{p}_0 - p_0|, |\hat{p}_1 - p_1|\} \leq (1 - p_0 - p_1)/4$ this implies

$$|\hat{b} - b| = 2 \cdot \max\{|\hat{p}_0 - p_0|, |\hat{p}_1 - p_1|\} \leq \frac{1}{2} \cdot (1 - p_0 - p_1) = \frac{b}{2},$$

which also implies $\hat{b} \geq b/2 > 0$. Hence, by (18) we deduce

$$\begin{aligned}|\hat{\eta}(x) - \eta(x)| &\leq \frac{4}{1 - p_0 - p_1} \cdot \max\{|(\hat{\eta}_{\text{corr}}(x) - \hat{p}_0) - (\eta_{\text{corr}}(x) - p_0)|, |(1 - \hat{p}_0 - \hat{p}_1) - (1 - p_0 - p_1)|\} \\ &\leq \frac{8}{1 - p_0 - p_1} \cdot \max\{|\hat{\eta}_{\text{corr}}(x) - \eta_{\text{corr}}(x)|, |\hat{p}_0 - p_0|, |\hat{p}_1 - p_1|\}.\end{aligned}$$

This completes the proof of Lemma 3.1. □