Activized Learning with Uniform Classification Noise: Supplementary Material

Liu Yang LIUY@CS.CMU.EDU

Machine Learning Department, Carnegie Mellon University

Steve Hanneke STEVE.HANNEKE@GMAIL.COM

This document provides specifications of the estimators used in Subroutine 1, along with a formal proof of Lemma 1.

1. Specification of Estimators

Following (Hanneke, 2009; 2012), we specify the estimators $\hat{\mathbb{P}}$ used in the algorithm as follows. For convenience, we suppose we have access to two independent sequences $W_1 = \{w_1, w_2, \ldots\}$ and $W_2 = \{w'_1, w'_2, \ldots\}$ of independent $\mathcal{D}$-distributed random variables, with (W_1, W_2) independent of $\mathcal{Z}$. Such sequences could easily be taken from the unlabeled data sequence in a preprocessing step, in which case we interpret the $\{X_i\}_{i=1}^{\infty}$ sequence referenced in the algorithms as those points remaining in the pool after extracting the sequences W_1 and W_2. Fix any $\mathcal{H} \subseteq \mathbb{C}$ and $m \in \mathbb{N}$. For any $k \in \mathbb{N}$, define

$$\mathcal{S}^{k-1}(\mathcal{H}) = \{S \in \mathcal{X}^{k-1} : \mathcal{H} \text{ shatters } S\}.$$

For any $(x, y) \in \mathcal{X} \times \{-1, +1\}$, define $\hat{\Gamma}_m^{(1)}(x, y, W_2, \mathcal{H}) = \mathbb{1}_{\bigcap_{h \in \mathcal{H}} \{h(x)\}}(y)$ and $\hat{\Delta}_m^{(1)}(x, W_2, \mathcal{H}) = \mathbb{1}_{\mathcal{S}^1(\mathcal{H})}(x)$. For any $k \in \{2, \ldots, d+1\}$, $\forall i \in \mathbb{N}$, let

$$S_i^{(k)} = \{w'_{1+(i-1)(k-1)}, \ldots, w'_{i(k-1)}\};$$

then let

$$M_m^{(k)}(\mathcal{H}) = \max\left\{1, \sum_{i=1}^{m^3} \mathbb{1}_{\mathcal{S}^{k-1}(\mathcal{H})}\left(S_i^{(k)}\right)\right\},$$

and for $(x, y) \in \mathcal{X} \times \{-1, +1\}$, define

$$\hat{\Gamma}_m^{(k)}(x, y, W_2, \mathcal{H}) = \frac{1}{M_m^{(k)}} \sum_{i=1}^{m^3} \mathbb{1}_{\mathcal{X}^{k-1} \setminus \mathcal{S}^{k-1}(\mathcal{H}[(x,-y)])}\left(S_i^{(k)}\right) \mathbb{1}_{\mathcal{S}^{k-1}(\mathcal{H})}\left(S_i^{(k)}\right),$$

$$\hat{\Delta}_m^{(k)}(x, W_2, \mathcal{H}) = \frac{1}{M_m^{(k)}(\mathcal{H})} \sum_{i=1}^{m^3} \mathbb{1}_{\mathcal{S}^k(\mathcal{H})}\left(S_i^{(k)} \cup \{x\}\right).$$

For any $k \in \{1, \ldots, d+1\}$, define

$$\hat{\Delta}_m^{(k)}(W_1, W_2, \mathcal{H}) = \frac{2}{m} + \frac{1}{m^3} \sum_{i=1}^{m^3} \mathbb{1}_{[1/8, \infty)}\left(\hat{\Delta}_m^{(k)}(w_i, W_2, \mathcal{H})\right).$$

With these definitions, we now define the estimators referenced in Subroutine 1 as follows. Letting n be the label budget argument to Subroutine 1, and letting m_n and V be as in Subroutine 1, for any $k \in \{1, \ldots, d+1\}$, $m \in \{m_n +$

$1, \ldots, n^{33/32}\}$, and $y \in \{-1, +1\}$, define

$$\hat{\mathbb{P}}\left(x : \hat{\mathbb{P}}\left(S \in \mathcal{X}^{k-1} : V \text{ shatters } S \cup \{x\} \middle| V \text{ shatters } S\right) \geq 1/2\right) = \hat{\Delta}_{m_n}^{(k)}(W_1, W_2, V),$$

$$\hat{\mathbb{P}}\left(S \in \mathcal{X}^{k-1} : V \text{ shatters } S \cup \{X_m\} \middle| V \text{ shatters } S\right) = \hat{\Delta}_m^{(k)}(X_m, W_2, V),$$

$$\hat{\mathbb{P}}\left(S \in \mathcal{X}^{k-1} : V[(X_m, -y)] \text{ does not shatter } S \middle| V \text{ shatters } S\right) = \hat{\Gamma}_m^{(k)}(X_m, y, W_2, V).$$

2. Relevant Lemmas

Before getting into the proof of Lemma 1, we first state a few relevant definitions and lemmas. The definitions are taken from (Hanneke, 2012), some of which are slightly modified here to suit our present context, and the lemmas represent slight generalizations of results proven by (Hanneke, 2012) (aside from the first lemma, which is due to (Vapnik and Chervonenkis, 1971)). Throughout this section, we fix an arbitrary distribution $\mathcal{D}_{XY} \in \mathrm{UniformNoise}(\mathbb{C})$.

For any $m \in \mathbb{N}$, let $V_m^\star = \{h \in \mathbb{C} : \forall i \leq m, h(X_m) = h_{\mathcal{D}_{XY}}^*(X_m)\}$. For any $\varepsilon > 0$, define $\mathrm{B}(h_{\mathcal{D}_{XY}}^*, \varepsilon) = \{h \in \mathbb{C} : \mathcal{D}(x : h(x) \neq h_{\mathcal{D}_{XY}}^*(x)) \leq \varepsilon\}$. For $k \in \mathbb{N} \cup \{0\}$ and $\mathcal{H} \subseteq \mathbb{C}$, let $\partial_{\mathcal{H}}^k h_{\mathcal{D}_{XY}}^* = \lim_{\varepsilon \to 0} \mathcal{S}^k(\mathcal{H} \cap \mathrm{B}(h_{\mathcal{D}_{XY}}^*, \varepsilon))$, and define

$$\tilde{d} = \min\left\{k \in \mathbb{N} : \mathcal{D}^k(\partial_{\mathbb{C}}^k h_{\mathcal{D}_{XY}}^*) = 0\right\},$$

where $\mathcal{D}^k$ is the product measure (with marginals equal $\mathcal{D}$) over $\mathcal{X}^k$. Also let $\tilde{\delta} = \mathcal{D}^{\tilde{d}-1}(\partial_{\mathbb{C}}^{\tilde{d}-1} h_{\mathcal{D}_{XY}}^*)$ (or $\tilde{\delta} = 1$ if $\tilde{d} = 1$). Additionally, for any $x \in \mathcal{X}$, $k \in \mathbb{N}$, and $\mathcal{H} \subseteq \mathbb{C}$, define $\mathring{p}_x(k, \mathcal{H}) = \mathcal{D}^{k-1}\left(S \in \mathcal{X}^{k-1} : S \cup \{x\} \in \mathcal{S}^k(\mathcal{H}) \middle| \mathcal{S}^{k-1}(\mathcal{H})\right)$ (recalling our convention that $\mathcal{X}^0 = \{\varnothing\}$, $\mathcal{D}^0(\mathcal{X}^0) = 1$, and $\varnothing \in \mathcal{S}^0(\mathcal{H})$ iff $\mathcal{H} \neq \{\}$). Here, and below, we adopt the simplifying convention that, for any probability measure P, and measurable sets A, B, if $P(B) = 0$, then $P(A|B) = 0$ by definition (when $P(B) > 0$, $P(A|B) = P(A \cap B)/P(B)$ as usual).

The first lemma is due to (Vapnik and Chervonenkis, 1971; Vapnik, 1982).

Lemma 4. *For any $\delta \in (0, 1)$ and $m \in \mathbb{N}$, on an event $E_m(\delta)$ of probability at least $1 - \delta/3$, every $h \in \mathbb{C}$ has*

$$|\operatorname{er}_m(h) - \operatorname{er}(h)| \leq \hat{U}_m(\delta)/2.$$

This immediately implies the following corollary.

Corollary 1. *If $\mathcal{D}_{XY} \in \mathrm{UniformNoise}(\mathbb{C})$, and Subroutine 1 is executed with label budget $n \in \mathbb{N}$ and confidence parameter $\delta \in (0, 1)$, then on the event $E_{m_n}(\delta)$,*

$$V_{m_n}^\star \subseteq V \subseteq \mathrm{B}\left(h_{\mathcal{D}_{XY}}^*, \phi(m_n; \delta)\right),$$

where $\phi(m_n; \delta) = \frac{2}{1 - 2\eta(\mathcal{D}_{XY})} \hat{U}_{m_n}(\delta)$.

Proof. On the event $E_{m_n}(\delta)$,

$$\operatorname{er}_{m_n}(h_{\mathcal{D}_{XY}}^*) - \min_{h' \in \mathbb{C}} \operatorname{er}_{m_n}(h') \leq \hat{U}_{m_n}(\delta)/2 + \operatorname{er}(h_{\mathcal{D}_{XY}}^*) - \min_{h' \in \mathbb{C}} \left(\operatorname{er}(h') - \hat{U}_{m_n}(\delta)/2\right) = \hat{U}_{m_n}(\delta),$$

and any $h \in \mathbb{C}$ with $\operatorname{er}_{m_n}(h) - \min_{h' \in \mathbb{C}} \operatorname{er}_{m_n}(h') \leq \hat{U}_{m_n}(\delta)$ has

$$\begin{aligned} \operatorname{er}(h) - \nu(\mathbb{C}; \mathcal{D}_{XY}) = \operatorname{er}(h) - \min_{h' \in \mathbb{C}} \operatorname{er}(h') &\leq \hat{U}_{m_n}(\delta)/2 + \operatorname{er}_{m_n}(h) - \min_{h' \in \mathbb{C}} (\operatorname{er}_{m_n}(h') - \hat{U}_{m_n}(\delta)/2) \\ &= \hat{U}_{m_n}(\delta) + \operatorname{er}_{m_n}(h) - \min_{h' \in \mathbb{C}} \operatorname{er}_{m_n}(h') \leq 2\hat{U}_{m_n}(\delta), \end{aligned}$$

so that $\mathcal{D}(x : h(x) \neq h_{\mathcal{D}_{XY}}^*(x)) = \frac{1}{1 - 2\eta(\mathcal{D}_{XY})}(\operatorname{er}(h) - \operatorname{er}(h_{\mathcal{D}_{XY}}^*)) \leq \frac{2}{1 - 2\eta(\mathcal{D}_{XY})} \hat{U}_{m_n}(\delta)$. □

The following lemmas are due to (Hanneke, 2012); some of them are slightly generalized to suit our present context, and we include proofs in those cases (which are essentially very minor modifications to the original proofs). The other lemmas are taken directly from (Hanneke, 2012), and the reader is referred to the original work of (Hanneke, 2012) for the proofs of these.

Lemma 5. *(Hanneke, 2012, part of Lemma 35) There is an event H' with $\mathbb{P}(H') = 1$ such that on H', $\forall m \in \mathbb{N}$, $\forall k \in \{0, \ldots, \tilde{d}-1\}$, for any $\mathcal{H}$ with $V_m^\star \subseteq \mathcal{H} \subseteq \mathbb{C}$,*

$$\mathcal{D}^k(\mathcal{S}^k(\mathcal{H})|\partial_{\mathbb{C}}^k h^*_{\mathcal{D}_{XY}}) = \mathcal{D}^k(\partial_{\mathcal{H}}^k h^*_{\mathcal{D}_{XY}}|\partial_{\mathbb{C}}^k h^*_{\mathcal{D}_{XY}}) = 1 \text{ and } \mathcal{D}^k(\partial_{\mathcal{H}}^k h^*_{\mathcal{D}_{XY}}) = \mathcal{D}^k(\partial_{\mathbb{C}}^k h^*_{\mathcal{D}_{XY}}).$$

Lemma 6. *(Hanneke, 2012, part of Lemma 36) There is a monotonic function $q(r) = o(1)$ (as $r \to 0$) such that, on the event H', for any $k \in \{0, \ldots, \tilde{d}-1\}$, $m \in \mathbb{N}$, $r > 0$, and set $\mathcal{H}$ such that $V_m^\star \subseteq \mathcal{H} \subseteq \mathrm{B}(h^*_{\mathcal{D}_{XY}}, r)$,*

$$\mathcal{D}^k(\mathcal{X}^k \setminus \partial_{\mathbb{C}}^k h^*_{\mathcal{D}_{XY}}|\mathcal{S}^k(\mathcal{H})) \leq q(r).$$

For any $\zeta \in (0,1)$, define $r_\zeta = \sup\{r \in (0,1) : q(r) < \zeta\}/2$; this value is guaranteed to exist, since $q(r) = o(1)$.

Lemma 7. *(based on Hanneke, 2012, Lemma 38) Fix any $\zeta \in (0,1)$ and $r \in [0, r_{\zeta/2}]$, and let $\mathring{\Delta}^{(\zeta)}(r) = (2/(\zeta\tilde{\delta}))\mathcal{D}^{\tilde{d}}(\mathcal{S}^{\tilde{d}}(\mathrm{B}(h^*_{\mathcal{D}_{XY}}, r)))$. On the event H', for any $m \in \mathbb{N}$ and any $\mathcal{H}$ with $V_m^\star \subseteq \mathcal{H} \subseteq \mathrm{B}(h^*_{\mathcal{D}_{XY}}, r)$,*

$$\mathcal{D}\left(x : \mathring{p}_x(\tilde{d}, \mathcal{H}) \geq \zeta\right) \leq \mathring{\Delta}^{(\zeta)}(r).$$

Proof. For any $x \in \mathcal{X}$,

$$\begin{aligned}
&\mathcal{D}^{\tilde{d}-1}\left(S \in \mathcal{X}^{\tilde{d}-1} : S \cup \{x\} \in \mathcal{S}^{\tilde{d}}(\mathcal{H}) \middle| \mathcal{S}^{\tilde{d}-1}(\mathcal{H})\right) \\
&= \mathcal{D}^{\tilde{d}-1}\left(S \in \mathcal{X}^{\tilde{d}-1} : S \cup \{x\} \in \mathcal{S}^{\tilde{d}}(\mathcal{H}) \middle| \mathcal{S}^{\tilde{d}-1}(\mathcal{H}) \cap \partial_{\mathbb{C}}^{\tilde{d}-1} h^*_{\mathcal{D}_{XY}}\right) \mathcal{D}^{\tilde{d}-1}\left(\partial_{\mathbb{C}}^{\tilde{d}-1} h^*_{\mathcal{D}_{XY}} \middle| \mathcal{S}^{\tilde{d}-1}(\mathcal{H})\right) \\
&\quad + \mathcal{D}^{\tilde{d}-1}\left(S \in \mathcal{X}^{\tilde{d}-1} : S \cup \{x\} \in \mathcal{S}^{\tilde{d}}(\mathcal{H}) \middle| \mathcal{S}^{\tilde{d}-1}(\mathcal{H}) \setminus \partial_{\mathbb{C}}^{\tilde{d}-1} h^*_{\mathcal{D}_{XY}}\right) \mathcal{D}^{\tilde{d}-1}\left(\mathcal{X}^{\tilde{d}-1} \setminus \partial_{\mathbb{C}}^{\tilde{d}-1} h^*_{\mathcal{D}_{XY}} \middle| \mathcal{S}^{\tilde{d}-1}(\mathcal{H})\right) \\
&\leq \mathcal{D}^{\tilde{d}-1}\left(S \in \mathcal{X}^{\tilde{d}-1} : S \cup \{x\} \in \mathcal{S}^{\tilde{d}}(\mathcal{H}) \middle| \mathcal{S}^{\tilde{d}-1}(\mathcal{H}) \cap \partial_{\mathbb{C}}^{\tilde{d}-1} h^*_{\mathcal{D}_{XY}}\right) + \mathcal{D}^{\tilde{d}-1}\left(\mathcal{X}^{\tilde{d}-1} \setminus \partial_{\mathbb{C}}^{\tilde{d}-1} h^*_{\mathcal{D}_{XY}} \middle| \mathcal{S}^{\tilde{d}-1}(\mathcal{H})\right). \qquad (1)
\end{aligned}$$

On the event H', Lemma 5 implies that the first term on the right hand side of (1) is equal

$$\begin{aligned}
&\mathcal{D}^{\tilde{d}-1}\left(S \in \mathcal{X}^{\tilde{d}-1} : S \cup \{x\} \in \mathcal{S}^{\tilde{d}}(\mathcal{H}) \middle| \mathcal{S}^{\tilde{d}-1}(\mathcal{H}) \cap \partial_{\mathbb{C}}^{\tilde{d}-1} h^*_{\mathcal{D}_{XY}}\right) \mathcal{D}^{\tilde{d}-1}\left(\mathcal{S}^{\tilde{d}-1}(\mathcal{H}) \middle| \partial_{\mathbb{C}}^{\tilde{d}-1} h^*_{\mathcal{D}_{XY}}\right) \\
&\qquad = \mathcal{D}^{\tilde{d}-1}\left(S \in \mathcal{X}^{\tilde{d}-1} : S \cup \{x\} \in \mathcal{S}^{\tilde{d}}(\mathcal{H}) \middle| \partial_{\mathbb{C}}^{\tilde{d}-1} h^*_{\mathcal{D}_{XY}}\right),
\end{aligned}$$

while Lemma 6 implies the second term on the right hand side of (1) is at most $q(r) < \zeta/2$, so that on H',

$$\begin{aligned}
&\mathcal{D}\left(x : \mathcal{D}^{\tilde{d}-1}\left(S \in \mathcal{X}^{\tilde{d}-1} : S \cup \{x\} \in \mathcal{S}^{\tilde{d}}(\mathcal{H}) \middle| \mathcal{S}^{\tilde{d}-1}(\mathcal{H})\right) \geq \zeta\right) \\
&\qquad \leq \mathcal{D}\left(x : \mathcal{D}^{\tilde{d}-1}\left(S \in \mathcal{X}^{\tilde{d}-1} : S \cup \{x\} \in \mathcal{S}^{\tilde{d}}(\mathcal{H}) \middle| \partial_{\mathbb{C}}^{\tilde{d}-1} h^*_{\mathcal{D}_{XY}}\right) > \zeta/2\right).
\end{aligned}$$

By Markov's inequality, this is at most

$$\frac{2}{\zeta\tilde{\delta}}\mathcal{D}^{\tilde{d}}\left(S \cup \{x\} \in \mathcal{X}^{\tilde{d}} : S \cup \{x\} \in \mathcal{S}^{\tilde{d}}(\mathrm{B}(h^*_{\mathcal{D}_{XY}}, r)) \text{ and } S \in \partial_{\mathbb{C}}^{\tilde{d}-1} h^*_{\mathcal{D}_{XY}}\right) \leq \mathring{\Delta}^{(\zeta)}(r). \qquad \square$$

Lemma 8. *(a special case of Hanneke, 2012, Lemma 41) There exist constants $\tau_1^* \in \mathbb{N}$ and $c_1, c_1' \in (0, \infty)$ such that, for any integer $\tau \geq \tau_1^*$, on an event $H_\tau^{(1)}$ with $\mathbb{P}(H_\tau^{(1)}) \geq 1 - c_1 \exp\{-c_1'\tau\}$, for any set $\mathcal{H}$ with $V_\tau^\star \subseteq \mathcal{H} \subseteq \mathrm{B}(h^*_{\mathcal{D}_{XY}}, r_{1/12})$, $\forall k \in \{1, \ldots, \tilde{d}\}$, for every integer $m > \tau$,*

$$\hat{\Delta}_m^{(k)}(X_m, W_2, \mathcal{H}) < 1/2 \implies \hat{\Gamma}_m^{(k)}(X_m, -h^*_{\mathcal{D}_{XY}}(X_m), W_2, \mathcal{H}) < \hat{\Gamma}_m^{(k)}(X_m, h^*_{\mathcal{D}_{XY}}(X_m), W_2, \mathcal{H}).$$

Lemma 9. *(based on Hanneke, 2012, Lemma 42) For any $\zeta \in (0,1)$, there exists a constant $c_2^{(\zeta)} \in (0, \infty)$ such that, for any $\tau \in \mathbb{N}$, and any set $\mathcal{H} \subseteq \mathbb{C}$ (possibly depending on $\mathcal{Z}$, but independent of W_2), on an event $H_\tau^{(2)}(\zeta)$ with $\mathbb{P}(H_\tau^{(2)}(\zeta)) \geq 1 - c_2^{(\zeta)} \cdot \exp\{-\zeta^2\tau^3\tilde{\delta}/8\}$, if $V_\tau^\star \subseteq \mathcal{H}$, then $\forall k \in \{1, \ldots, \tilde{d}\}$, for every integer $m \geq \tau$,*

$$\mathcal{D}\left(x : \left|\mathring{p}_x(k, \mathcal{H}) - \hat{\Delta}_m^{(k)}(x, W_2, \mathcal{H})\right| > \zeta\right) \leq \exp\{-\zeta^2 m^3\tilde{\delta}/2\}.$$

The proof of Lemma 9 is identical to that of (Hanneke, 2012, Lemma 42), except substituting $\mathcal{H} \cup V_\ell^\star$ for $V_\ell^\star$ throughout that proof (for each $\ell \geq \tau$) and in the definitions (p_x, $\hat{p}_x$, $\tilde{m}$) referenced therein; the proof remains valid with this modification. For brevity, we do not repeat the details here.

Lemma 10. *(based on Hanneke, 2012, Lemma 43) For any $\alpha, \zeta \in (0,1)$, $\beta \in (0, 1-\sqrt{\alpha}]$, $\tau \in \mathbb{N}$, and any set $\mathcal{H} \subseteq \mathbb{C}$ (possibly depending on $\mathcal{Z}$, but independent of W_2), on the event $H_\tau^{(2)}(\beta\zeta)$ (defined relative to this $\mathcal{H}$), if $V_\tau^\star \subseteq \mathcal{H}$, then $\forall k \in \{1, \ldots, \tilde{d}\}$, for every integer $m \geq \tau$,*

$$\mathcal{D}\left(x : \hat{\Delta}_m^{(k)}(x, W_2, \mathcal{H}) \geq \zeta\right) \leq \mathcal{D}\left(x : \mathring{p}_x(k, \mathcal{H}) \geq \sqrt{\alpha}\zeta\right) + \exp\{-\beta^2\zeta^2 m^3 \tilde{\delta}/2\}.$$

Proof. Fix $\alpha, \zeta, \beta, \tau, \mathcal{H}$ as in the lemma statement, and suppose the event $H_\tau^{(2)}(\beta\zeta)$ occurs; also suppose $V_\tau^\star \subseteq \mathcal{H}$. Fix any $k \in \{1, \ldots, \tilde{d}\}$ and $m \in \mathbb{N}$ with $m \geq \tau$. By a union bound,

$$\mathcal{D}\left(x : \hat{\Delta}_m^{(k)}(x, W_2, \mathcal{H}) \geq \zeta\right) \leq \mathcal{D}\left(x : \mathring{p}_x(k, \mathcal{H}) \geq \sqrt{\alpha}\zeta\right) + \mathcal{D}\left(x : \left|\mathring{p}_x(k, \mathcal{H}) - \hat{\Delta}_m^{(k)}(x, W_2, \mathcal{H})\right| > (1-\sqrt{\alpha})\zeta\right).$$

Since $(1-\sqrt{\alpha})\zeta \geq \beta\zeta$, Lemma 9 and monotonicity of probability measures imply

$$\mathcal{D}\left(x : \left|\mathring{p}_x(k, \mathcal{H}) - \hat{\Delta}_m^{(k)}(x, W_2, \mathcal{H})\right| > (1-\sqrt{\alpha})\zeta\right) \leq \exp\{-\beta^2\zeta^2 m^3 \tilde{\delta}/2\}. \qquad \square$$

Lemma 11. *(based on Hanneke, 2012, Lemma 44) For any $\xi \in (0, 1/32]$, there is a constant $\tau_2^*(\xi) \in \mathbb{N}$ such that, for any set $\mathcal{H} \subseteq \mathbb{C}$ (possibly depending on $\mathcal{Z}$, but independent of (W_1, W_2)), and any integer $\tau \geq \tau_2^*(\xi)$, there is an event $H_\tau^{(3)}$ with $\mathbb{P}(H_\tau^{(3)}) \geq 1 - 2\tilde{d} \cdot \exp\{-2\tau\}$ such that, on $H_\tau^{(3)} \cap H_\tau^{(2)}(\xi)$ (where $H_\tau^{(2)}(\xi)$ is defined relative to this $\mathcal{H}$), if $V_\tau^\star \subseteq \mathcal{H}$, then $\forall k \in \{1, \ldots, \tilde{d}\}$,*

$$\mathcal{D}\left(x : \mathring{p}_x(k, \mathcal{H}) \geq 1/4\right) + \exp\{-\tau^3\tilde{\delta}/2048\} \leq \hat{\Delta}_\tau^{(k)}(W_1, W_2, \mathcal{H}) \leq \mathcal{D}\left(x : \mathring{p}_x(k, \mathcal{H}) \geq 1/16\right) + 4\tau^{-1}.$$

Proof. Fix any $\xi \in (0, 1/32]$, define $\tau_2^*(\xi) = \left\lceil \left(\frac{4}{\tilde{\delta}\xi^2} \ln\left(\frac{4}{\tilde{\delta}\xi^2}\right)\right)^{1/3} \right\rceil$, and fix any integer $\tau \geq \tau_2^*(\xi)$. For any $k \in \{1, \ldots, \tilde{d}\}$, Hoeffding's inequality and the law of total probability imply that, on an event $G_\tau(k)$ with $\mathbb{P}(G_\tau(k)) \geq 1 - 2\exp\{-2\tau\}$, we have

$$\left|\mathcal{D}\left(x : \hat{\Delta}_\tau^{(k)}(x, W_2, \mathcal{H}) \geq 1/8\right) - \tau^{-3} \sum_{i=1}^{\tau^3} \mathbb{1}_{[1/8, \infty)}\left(\hat{\Delta}_\tau^{(k)}(w_i, W_2, \mathcal{H})\right)\right| \leq \tau^{-1}. \tag{2}$$

Defining the event $H_\tau^{(3)} = \bigcap_{k=1}^{\tilde{d}} G(k)$ yields $\mathbb{P}(H_\tau^{(3)}) \geq 1 - 2\tilde{d} \cdot \exp\{-2\tau\}$ by a union bound.

Now fix any $k \in \{1, \ldots, \tilde{d}\}$. By a union bound,

$$\mathcal{D}\left(x : \mathring{p}_x(k, \mathcal{H}) \geq 1/4\right) \leq \mathcal{D}\left(x : \hat{\Delta}_\tau^{(k)}(x, W_2, \mathcal{H}) \geq 1/8\right) + \mathcal{D}\left(x : \left|\mathring{p}_x(k, \mathcal{H}) - \hat{\Delta}_\tau^{(k)}(x, W_2, \mathcal{H})\right| > 1/8\right). \tag{3}$$

By Lemma 9, on $H_\tau^{(2)}(\xi)$, if $V_\tau^\star \subseteq \mathcal{H}$, then

$$\mathcal{D}\left(x : \left|\mathring{p}_x(k, \mathcal{H}) - \hat{\Delta}_\tau^{(k)}(x, W_2, \mathcal{H})\right| > 1/8\right) \leq \mathcal{D}\left(x : \left|\mathring{p}_x(k, \mathcal{H}) - \hat{\Delta}_\tau^{(k)}(x, W_2, \mathcal{H})\right| > \xi\right) \leq \exp\{-\xi^2\tau^3\tilde{\delta}/2\}. \tag{4}$$

Furthermore, on the event $H_\tau^{(3)}$, (2) implies

$$\mathcal{D}\left(x : \hat{\Delta}_\tau^{(k)}(x, W_2, \mathcal{H}) \geq 1/8\right) \leq \hat{\Delta}_\tau^{(k)}(W_1, W_2, \mathcal{H}) - \tau^{-1}.$$

Combining this with (3) and (4) reveals

$$\mathcal{D}\left(x : \mathring{p}_x(k, \mathcal{H}) \geq 1/4\right) \leq \hat{\Delta}_\tau^{(k)}(W_1, W_2, \mathcal{H}) - \tau^{-1} + \exp\{-\xi^2\tau^3\tilde{\delta}/2\} \leq \hat{\Delta}_\tau^{(k)}(W_1, W_2, \mathcal{H}) - \exp\{-\tau^3\tilde{\delta}/2048\},$$

where this last inequality is due to our choice of $\tau_2^*(\xi)$. In particular, we have established the first inequality in the statement of the lemma.

Toward establishing the second inequality, note that on $H_\tau^{(3)}$, (2) implies

$$\hat{\Delta}_\tau^{(k)}(W_1, W_2, \mathcal{H}) \le \mathcal{D}\left(x : \hat{\Delta}_\tau^{(k)}(x, W_2, \mathcal{H}) \ge 1/8\right) + 3\tau^{-1},$$

while Lemma 10 (with $\alpha = 1/4$, $\zeta = 1/8$, $\beta = \xi/\zeta < 1 - \sqrt{\alpha}$) implies that on $H_\tau^{(2)}(\xi)$, if $V_\tau^\star \subseteq \mathcal{H}$, then

$$\mathcal{D}\left(x : \hat{\Delta}_\tau^{(k)}(x, W_2, \mathcal{H}) \ge 1/8\right) \le \mathcal{D}\left(x : \mathring{p}_x(k, \mathcal{H}) \ge 1/16\right) + \exp\{-\xi^2\tau^3\tilde{\delta}/2\}.$$

Combining these two inequalities, we have

$$\hat{\Delta}_\tau^{(k)}(W_1, W_2, \mathcal{H}) \le \mathcal{D}\left(x : \mathring{p}_x(k, \mathcal{H}) \ge 1/16\right) + \exp\{-\xi^2\tau^3\tilde{\delta}/2\} + 3\tau^{-1}.$$

Noting that our choice of $\tau_2^*(\xi)$ guarantees $\exp\{-\xi^2\tau^3\tilde{\delta}/2\} \le \tau^{-1}$, this establishes the second inequality in the lemma statement. $\square$

3. Proof of Lemma 1

We are now ready for the proof of Lemma 1. For each $k \in \{1, \ldots, d+1\}$, and $n \in \mathbb{N}$, let

$$\mathcal{U}_n^{(k)} = \left\{m_n + 1, \ldots, m_n + \left\lfloor n/\left(6 \cdot 2^k \hat{\Delta}_{m_n}^{(k)}(W_1, W_2, V)\right)\right\rfloor\right\},$$

where m_n and V are as in Subroutine 1, when executed with label budget n and confidence parameter δ_n.

Proof of Lemma 1. This proof follows closely to the proof of (Hanneke, 2012, Lemma 45). Let $n^* = \max\left\{\left(\frac{64}{\min\{r_{1/12}, r_{1/32}\}(1-2\eta(\mathcal{D}_{XY}))}\right)^4, \left(\frac{64}{\min\{r_{1/12}, r_{1/32}\}(1-2\eta(\mathcal{D}_{XY}))}\right)^2\left(\ln(12) + d\ln\left(\frac{32e}{\min\{r_{1/12}, r_{1/32}\}(1-2\eta(\mathcal{D}_{XY}))}\right)\right),\right.$ $\left. 2\tau_1^*,\ 2\tau_2^*(1/32)\right\}$. By setting c appropriately large, we can guarantee the result is trivially satisfied for all $n < n^*$ (e.g., by taking $c \ge \exp\{c'(n^*)^{1/3}\}$). Now fix any integer $n \ge n^*$, and let V and m_n be as in Subroutine 1, when executed with label budget n and confidence parameter δ_n. In particular, note that $m_n \ge \max\{\tau_1^*, \tau_2^*(1/32)\}$, and $\phi(m_n, \delta_n) \le \min\{r_{1/12}, r_{1/32}\}$.

Let $H_{m_n}^{(2)}(1/32)$ and $H_{m_n}^{(3)}$ be the events from Lemmas 9 and 11, respectively, defined relative to the set V. By Corollary 1, and Lemmas 10 and 11, on the event $E_{m_n}(\delta_n) \cap H_{m_n}^{(2)}(1/32) \cap H_{m_n}^{(3)}$, $\forall k \in \{1, \ldots, \tilde{d}\}$, $\forall m \in \mathcal{U}_n^{(k)}$,

$$\begin{aligned}\mathcal{D}\left(x : \hat{\Delta}_m^{(k)}(x, W_2, V) \ge 1/2\right) &\le \mathcal{D}\left(x : \mathring{p}_x(k, V) \ge 1/4\right) + \exp\left\{-m^3\tilde{\delta}/2048\right\} \\ &\le \mathcal{D}\left(x : \mathring{p}_x(k, V) \ge 1/4\right) + \exp\left\{-m_n^3\tilde{\delta}/2048\right\} \le \hat{\Delta}_{m_n}^{(k)}(W_1, W_2, V). \qquad (5)\end{aligned}$$

Noting that, $\forall k \in \{1, \ldots, \tilde{d}\}$, $\{X_m : m \in \mathcal{U}_n^{(k)}\}$ are conditionally independent given $\hat{\Delta}_{m_n}^{(k)}(W_1, W_2, V)$, with each of these X_m variables having conditional distribution $\mathcal{D}$, a Chernoff bound implies that (on the above event),

$$\begin{aligned}&\mathbb{P}\left(\left|\left\{m \in \mathcal{U}_n^{(k)} : \hat{\Delta}_m^{(k)}(X_m, W_2, V) \ge 1/2\right\}\right| > n/\left(3 \cdot 2^k\right) \middle| W_1, W_2, V\right) \\ &\le \mathbb{P}\left(\left|\left\{m \in \mathcal{U}_n^{(k)} : \hat{\Delta}_m^{(k)}(X_m, W_2, V) \ge 1/2\right\}\right| > 2\left|\mathcal{U}_n^{(k)}\right| \hat{\Delta}_{m_n}^{(k)}(W_1, W_2, V) \middle| W_1, W_2, V\right) \\ &\le \exp\left\{-\left\lfloor n/\left(6 \cdot 2^k \hat{\Delta}_{m_n}^{(k)}(W_1, W_2, V)\right)\right\rfloor \hat{\Delta}_{m_n}^{(k)}(W_1, W_2, V)/3\right\} \le \exp\left\{1 - n/\left(18 \cdot 2^k\right)\right\}.\end{aligned}$$

By the law of total probability and a union bound, there exists an event $H_{m_n}^{(4)}$ with

$$\mathbb{P}(E_{m_n}(\delta_n) \cap H_{m_n}^{(2)}(1/32) \cap H_{m_n}^{(3)} \setminus H_{m_n}^{(4)}) \le \tilde{d} \cdot \exp\left\{1 - n/\left(18 \cdot 2^{\tilde{d}}\right)\right\}$$

such that on $E_{m_n}(\delta_n) \cap H_{m_n}^{(2)}(1/32) \cap H_{m_n}^{(3)} \cap H_{m_n}^{(4)}$, $\forall k \in \{1, \ldots, \tilde{d}\}$,

$$\left|\left\{m \in \mathcal{U}_n^{(k)} : \hat{\Delta}_m^{(k)}(X_m, W_2, V) \ge 1/2\right\}\right| \le \left\lfloor n/\left(3 \cdot 2^k\right)\right\rfloor.$$

In particular, this implies that on this event, the "$t < n$" condition in Step 6 of Subroutine 1 is redundant while $k \leq \tilde{d}$, so that every time the algorithm reaches Step 9 while $k \leq \tilde{d}$, we have $\hat{\Delta}_m^{(k)}(X_m, W_2, V) < 1/2$. Combined with Lemma 8, this implies that on the event $E_{m_n}(\delta_n) \cap H_{m_n}^{(1)} \cap H_{m_n}^{(2)}(1/32) \cap H_{m_n}^{(3)} \cap H_{m_n}^{(4)}$, every value of m for which the algorithm reaches Step 9 while $k \leq \tilde{d}$ has $\hat{\Gamma}_m^{(k)}(X_m, -h^*_{\mathcal{D}_{XY}}(X_m), W_2, V) < \hat{\Gamma}_m^{(k)}(X_m, h^*_{\mathcal{D}_{XY}}(X_m), W_2, V)$, so that $\hat{y} = h^*_{\mathcal{D}_{XY}}(X_m)$. In particular, this means that $\mathrm{er}_{\mathcal{L}_k}(h^*_{\mathcal{D}_{XY}}) = 0$ for every $k \leq \tilde{d}$.

Finally, by Lemmas 11 and 7, on the event $E_{m_n}(\delta_n) \cap H' \cap H_{m_n}^{(2)}(1/32) \cap H_{m_n}^{(3)}$, we have

$$\hat{\Delta}_{m_n}^{(\tilde{d})}(W_1, W_2, V) \leq \mathcal{D}\left(x : \mathring{p}_x(\tilde{d}, V) \geq 1/16\right) + 4m_n^{-1} \leq \mathring{\Delta}^{(1/16)}(\phi(m_n; \delta_n)) + 4m_n^{-1}.$$

Thus, if we define $\phi_1(n) = \min\left\{\left\lfloor n/\left(6 \cdot 2^{\tilde{d}}\left(\mathring{\Delta}^{(1/16)}(\phi(m_n;\delta_n)) + 4m_n^{-1}\right)\right)\right\rfloor, n^{33/32}\right\}$, we have that on the event $E_{m_n}(\delta_n) \cap H' \cap H_{m_n}^{(2)}(1/32) \cap H_{m_n}^{(3)}$, $|\mathcal{L}_{\tilde{d}} \cup Q_{\tilde{d}}| \geq \phi_1(n)$. Since $\mathcal{D}^{\tilde{d}}(\partial_{\mathbb{C}}^{\tilde{d}} h^*_{\mathcal{D}_{XY}}) = 0$, continuity of probability measures implies $\mathring{\Delta}^{(1/16)}(r) \to 0$ as $r \to 0$; furthermore, since $\phi(m_n; \delta_n) \to 0$ as $n \to \infty$, we have $\mathring{\Delta}^{(1/16)}(\phi(m_n; \delta_n)) \to 0$ as $n \to \infty$. We also have $4m_n^{-1} \to 0$ as $n \to \infty$. Altogether, this implies $\phi_1(n) = \omega(n)$.

Thus, the requirements on $\mathcal{L}_{k^*}$ and Q_{k^*} in Lemma 1 are satisfied for $k^* = \tilde{d}$, on the event $E_{m_n}(\delta_n) \cap H' \cap H_{m_n}^{(1)} \cap H_{m_n}^{(2)}(1/32) \cap H_{m_n}^{(3)} \cap H_{m_n}^{(4)}$, which has (by a union bound)

$$\begin{aligned}
&\mathbb{P}\left(E_{m_n}(\delta_n) \cap H' \cap H_{m_n}^{(1)} \cap H_{m_n}^{(2)}(1/32) \cap H_{m_n}^{(3)} \cap H_{m_n}^{(4)}\right) \\
&\geq 1 - (1 - \mathbb{P}(E_{m_n}(\delta_n))) - (1 - \mathbb{P}(H')) - (1 - \mathbb{P}(H_{m_n}^{(1)})) - (1 - \mathbb{P}(H_{m_n}^{(2)}(1/32))) - (1 - \mathbb{P}(H_{m_n}^{(3)})) \\
&\qquad - \mathbb{P}(E_{m_n}(\delta_n) \cap H_{m_n}^{(2)}(1/32) \cap H_{m_n}^{(3)} \setminus H_{m_n}^{(4)}) \\
&\geq 1 - (1/3)\exp\{-\sqrt{n}\} - c_1 \exp\{-c_1' m_n\} - c_2^{(1/32)} \exp\{-m_n^3 \tilde{\delta}/2^{13}\} - 2\tilde{d}\exp\{-2m_n\} - \tilde{d}\exp\left\{1 - n/\left(18 \cdot 2^{\tilde{d}}\right)\right\} \\
&\geq 1 - (4 + c_1 + c_2^{(1/32)} + 5\tilde{d}) \cdot \exp\left\{-\min\left\{c_1'/4, \tilde{\delta}/2^{14}, 1/(18 \cdot 2^{\tilde{d}})\right\}\sqrt{n}\right\}.
\end{aligned}$$

The result therefore holds by taking $c' = \min\left\{c_1'/4, \tilde{\delta}/2^{14}, 1/(18 \cdot 2^{\tilde{d}})\right\}$ and $c = (4 + c_1 + c_2^{(1/32)} + 5\tilde{d} + \exp\{c'(n^*)^{1/2}\})$; (we have in fact proven a slightly stronger result than stated in Lemma 1, increasing $n^{1/3}$ to $n^{1/2}$). □

References

Hanneke, S. *Theoretical Foundations of Active Learning*. PhD thesis, Machine Learning Department, School of Computer Science, Carnegie Mellon University, 2009.

Hanneke, S. Activized learning: Transforming passive to active with improved label complexity. *Journal of Machine Learning Research*, 13(5):1469–1587, 2012.

Vapnik, V. *Estimation of Dependencies Based on Empirical Data*. Springer-Verlag, New York, 1982.

Vapnik, V. and Chervonenkis, A. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16:264–280, 1971.