A Trichotomy for Transductive Online Learning
Supplementary Materials

A Multiclass Threshold Bounds

Definition A.1. *Let $\mathcal{X}$ and $\mathcal{Y}$ be sets, let $X = \{x_1, \dots, x_t\} \subseteq \mathcal{X}$, and let $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$. We say that X is threshold-shattered by $\mathcal{H}$ if there exist distinct $y_0, y_1 \in \mathcal{Y}$ and functions $h_1, \dots, h_t \in \mathcal{H}$ such that $h_i(x_j) = y_{\mathbb{1}(j \leq i)}$. The threshold dimension of $\mathcal{H}$, denoted $\mathsf{TD}(\mathcal{H})$, is the supremum of the set of integers t for which there exists a threshold-shattered set of cardinality t.*

We introduce the following generalization of the threshold dimension.

Definition A.2. *Let $\mathcal{X}$ and $\mathcal{Y}$ be sets, let $X = \{x_1, \dots, x_t\} \subseteq \mathcal{X}$, and let $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$. We say that X is multi-class threshold-shattered by $\mathcal{H}$ if there exist $y_1, y'_1 \dots, y_t, y'_t \in \mathcal{Y}$ such that $y_i \neq y'_j$ for all $i, j \in [t]$, and there exist functions $h_1, \dots, h_t \in \mathcal{H}$ such that*

$$h_i(x_j) = \begin{cases} y_i & (j \leq i) \\ y'_j & (j > i). \end{cases}$$

The multi-class threshold dimension of $\mathcal{H}$, denoted $\mathsf{MTD}(\mathcal{H})$, is the supremum of the set of integers t for which there exists a threshold-shattered set of cardinality t.

Claim A.3. *Let $\mathcal{X}$ and $\mathcal{Y}$ be sets, $k = |\mathcal{Y}| < \infty$, and let $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$. Then $\mathsf{TD}(\mathcal{H}) \geq \lfloor \mathsf{MTD}(\mathcal{H})/k^2 \rfloor$.*

Proof of Claim A.3. The proof follows from two applications of the pigeonhole principle. □

Claim A.4. *Let $\mathcal{X}$ and $\mathcal{Y}$ be sets, let $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$ such that $d = \mathsf{TD}(\mathcal{H}) < \infty$, and let $n \in \mathbb{N}$. Then*

$$M(\mathcal{H}, n) \geq \min \{\lfloor \log(d) \rfloor, \lfloor \log(n) \rfloor\}.$$

The proof of Claim A.4 is similar to that of Claim 3.4.

Theorem A.5. *Let $\mathcal{X}$ and $\mathcal{Y}$ be sets with $k = |\mathcal{Y}| < \infty$, let $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$. If $\mathsf{LD}(\mathcal{H}) = \infty$ then $\mathsf{MTD}(\mathcal{H}) = \infty$.*

Following is a lemma from Ramsey theory used for proving Theorem A.5, and a generalized notion of subtrees used in that lemma.

Definition A.6. *Let X be a finite set and let $(X, \preceq)$ be a partial order relation. For $p, c \in X$, we say that c is a child of p if $p \preceq c$ and there does not exist $m \in X$ such that $p \preceq m \preceq c$. We say that $z \in X$ is a leaf if there exists no $x \in X$ such that $z \preceq x$. $(X, \preceq)$ is a binary tree every non-leaf $x \in X$ has precisely 2 children. The depth of $z \in X$ is the largest $d \in \mathbb{N}$ for which there exist distinct $x_1, \dots, x_d \in X$ such that $x_1 \preceq x_2 \preceq \cdots \preceq x_d \preceq z$. For $d \in \mathbb{N}$, we say that $(X, \preceq)$ is a complete binary tree of depth d if $(X, \preceq)$ is a binary tree and all the leaves in X have depth d. We say that a partial order $(X', \preceq')$ is a subtree of $(X, \preceq)$ if $X' \subseteq X$, and $\forall a, b \in X' : \; a \preceq' b \iff a \preceq b$.*

The following lemma follows from Lemma 16 in Appendix B of [ALMM19].

Lemma A.7. *Let $k, d \in \mathbb{N}$, and let $\mathcal{Y}$ be a set, $|\mathcal{Y}| = k$. Let $T = (X, \preceq)$ be a complete binary tree of depth $d \in \mathbb{N}$, and let $g : \; X \to \mathcal{Y}$. Then T has a monochromatic complete binary tree subtree $T' = (X', \preceq')$ of depth d/k, namely there exists T' such that T' is a subtree of T, T' is a complete binary tree of depth d/k, and $|g(X')| = |\{g(a) : a \in X'\}| = 1$.*

Proof of Theorem A.5. Let $f_k(d)$ be the largest number such that every class with Littlestone dimension d has multi-class threshold dimension at least $f_k(d)$. We show by induction on d that f_k satisfies the following recurrence relation: $f_k(d) \geq 1 + f_k(\lceil d/k \rceil - 1)$.

For the base case, if $d = \mathsf{LD}(\mathcal{H}) = 0$, $\mathcal{H}$ and $\mathcal{X}$ are non-empty and therefore $\mathsf{MTD}(\mathcal{H}) \geq 1$. For the induction step $d = \mathsf{LD}(\mathcal{H}) \geq 1$, let T be a Littlestone tree of depth d that is shattered by $\mathcal{H}$. Let $h \in \mathcal{H}$. Then h is a k-cloring of the nodes of T. By Lemma A.7, there exists an h-monochromatic subtree $T' \subseteq T$ of depth at least d/k. Let y_1 be the color assigned by h to all nodes of T'. T' is shattered by $\mathcal{H}$, so there exists a child x_1 of the root r of T' such that the label of the edge leading to it is some $y_1' \neq y_1$. Let $\mathcal{H}_1 = \{h \in \mathcal{H} : \ h(x_1) = y_1'\}$. Notice that $\mathsf{LD}(\mathcal{H}_1) \geq d/k - 1$, so by the induction hypothesis, there exist $x_2, \ldots, x_s$ for $s = f_k(\lceil d/k \rceil - 1)$ that are multi-class threshold shattered. By construction, the set $\{x_1, \ldots, x_s\}$ is multi-class threshold shattered by $\mathcal{H}$, as desired. □

B Multiclass Trichotomy

The Natarajan dimension is one popular generalization of the VC dimension to the multiclass setting.

Definition B.1 ([Nat89]). *Let $\mathcal{X}$ and $\mathcal{Y}$ be sets, let $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$, let $d \in \mathbb{N}$, and let $X = \{x_1, \ldots, x_d\} \subseteq \mathcal{X}$. We say that $\mathcal{H}$ Natarajan-shatters X if there exist $f_0, f_1 : X \to \mathcal{Y}$ such that:*

1. *$\forall x \in X : \ f_0(x) \neq f_1(x)$; and*
2. *$\forall A \subseteq X \ \exists h \in \mathcal{H} \ \forall x \in X : \ h(x) = f_{\mathbb{1}(x \in A)}(x)$.*

The Natarajan dimension of $\mathcal{H}$ is $\mathsf{ND}(\mathcal{H}) = \sup\{|X| : \ X \subseteq \mathcal{X}$ finite $\wedge$ $\mathcal{H}$ Natarajan-shatters $X\}$.

We show the following generalization of Theorem 4.1 for the multiclass setting.

Theorem B.2 (Formal Version of Theorem 5.1). *Let $\mathcal{X}$ and $\mathcal{Y}$ be sets with $k = |\mathcal{Y}| < \infty$, let $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$, and let $n \in \mathbb{N}$ such that $n \leq |\mathcal{X}|$.*

1. *If $\mathsf{ND}(\mathcal{H}) = \infty$ then $M(\mathcal{H}, n) = n$.*
2. *Otherwise, if $\mathsf{ND}(\mathcal{H}) = d < \infty$ and $\mathsf{LD}(\mathcal{H}) = \infty$ then*
$$\max\{\min\{d, n\}, \lfloor \log(n) \rfloor\} \leq M(\mathcal{H}, n) \leq O(d \log(nk/d)). \tag{5}$$
The $\Omega(\cdot)$ and $O(\cdot)$ notations hide universal constants that do not depend on $\mathcal{X}$, $\mathcal{Y}$ or $\mathcal{H}$.
3. *Otherwise, there exists a number $C(\mathcal{H}) \in \mathbb{N}$ (that depends on $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{H}$ but does not depend on n) such that $M(\mathcal{H}, n) \leq C(\mathcal{H})$.*

The proof of Theorem B.2 uses the following generalization of the Sauer–Shelah–Perles lemma.

Theorem B.3 ([Nat89]; Corollary 5 in [HL95]). *Let $d, n, k \in \mathbb{N}$, let $\mathcal{X}$ and $\mathcal{Y}$ be sets of cardinality n and k respectively, and let $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$ such that $\mathsf{ND}(\mathcal{H}) \leq d$. Then*

$$|\mathcal{H}| \leq \sum_{i=0}^{d} \binom{n}{i} \binom{k+1}{2}^i \leq \left(\frac{enk^2}{d}\right)^d.$$

Proof of Theorem B.2. Items 1 and 3 and the $\min\{d, n\}$ lower bound in Item 2 follow similarly to the corresponding items in Theorem 4.1. The upper bound in Item 2 also follows similarly to the corresponding item in Theorem 4.1, except that it uses Theorem B.3 instead of the Sauer–Shelah–Perles lemma.

The $\lfloor \log(n) \rfloor$ lower bound in Item 2 follows from Theorem A.5 and Claim A.4. □