Quiz: Chapter 1+2 definitions

Definition 1. Let $\mathbf{x}_1, ..., \mathbf{x}_N \in \mathcal{X}$. The *dichotomies* generated by a hypothesis class $\mathcal{H}$ on these points are defined by

$$\mathcal{H}(\mathbf{x}_1, ..., \mathbf{x}_N) = \Big\{ \big(h(\mathbf{x}_1), ..., h(\mathbf{x}_N)\big) : h \in \mathcal{H} \Big\}$$

Definition 2. The *growth function* for a hypothesis class $\mathcal{H}$ is defined to be

$$m_{\mathcal{H}}(N) = \max_{\mathbf{x}_1, ..., \mathbf{x}_N \in \mathcal{X}} \big|\mathcal{H}(\mathbf{x}_1, ..., \mathbf{x}_N)\big|.$$

Definition 3. We say that a hypothesis class $\mathcal{H}$ can *shatter* a dataset $\mathbf{x}_1, ..., \mathbf{x}_N$ if any of the following equivalent statements are true:

1. $\mathcal{H}$ is capable of generating all possible dichotomies of $\mathbf{x}_1, ..., \mathbf{x}_N$.
2. $\mathcal{H}(\mathbf{x}_1, ..., \mathbf{x}_N) = \{-1, +1\}^N$.
3. $|\mathcal{H}(\mathbf{x}_1, ..., \mathbf{x}_N)| = 2^N$.

NOTE: You must list all 3 statements.

Definition 4. The integer k is said to be a *break point* for hypothesis class $\mathcal{H}$ if

no data set of size k can be shattered by $\mathcal{H}$.

Definition 5. The *Vapnik-Chervonenkis dimension* (VC dimension) of a hypothesis class $\mathcal{H}$, denoted by $d_{\mathrm{VC}}(\mathcal{H})$ or simply d_{VC}, is

> the largest value of N for which $m_{\mathcal{H}}(N) = 2^N$. If $m_{\mathcal{H}}(N) = 2^N$ for all N, then $d_{\mathrm{VC}} = \infty$.

Theorem 1 (VC generalization bound)**.** For any tolerance $\delta > 0$, we have that with probability at least $1 - \delta$,

$$E_{\mathrm{out}} \leq E_{\mathrm{in}} + O\left(\sqrt{\frac{d_{\mathrm{VC}} \log N - \log \delta}{N}}\right).$$

NOTE: The more precise, non-asymptotic formulas would also be acceptable.

Theorem 2 (Finite Hypothesis Class Generalization Theorem)**.** For any tolerance $\delta > 0$, we have that with probability at least $1 - \delta$,

$$E_{\mathrm{out}}(g) \leq E_{\mathrm{in}}(g) + \sqrt{\frac{\log 2M - \log \delta}{2N}}.$$

NOTE: The less precise, asymptotic formulas would also be acceptable.

Theorem 3 (Hoeffding Inequality)**.** Let $a_1, ..., a_N$ be N independent and identically distributed random variables satisfying $0 \leq a_i \leq 1$. Let $\nu = \frac{1}{n}\sum_{i=1}^{N} a_i$ be the empirical average and $\mu = \mathbb{E}\nu$ be the true mean of the underlying distribution.

Then, for all $\epsilon > 0$,

$$\mathbb{P}\left(|\nu - \mu| \geq \epsilon\right) \leq 2\exp(-2\epsilon^2 N). \tag{1}$$

NOTE: The version based on $1 - \delta$ would also be acceptable. Asymptotic versions of either inequality are also acceptable.