# Problem Set 5

Instructions:

- This problem set covers Unit 10 – Nonuniform Learning. The textbook reading corresponding to this unit is Chapters 7 and 31 in SSBD. You may cite without proof any claim that was proved in the lectures or in the textbook reading material.
- Before you start, make sure you are familiar with the course's Homework Policy.

**1.** Let $\mathcal{X}$ be a set, let $\mathcal{H}$ be a finite class of functions $\mathcal{X} \to \{0,1\}$, and let $d : \mathcal{H} \to \{0,1\}^*$ be an injective function (an encoding of $\mathcal{H}$). For any bit string $w \in \{0,1\}^*$, let $|w|$ denote the the length of $w$.

**(a)** Show that $\mathsf{VC}(\mathcal{H}) \leq 2 \cdot \max_{h\in\mathcal{H}} |d(h)|$. [10 pts]

**(b)** We say that $d$ is *prefix-free* if for any $h, h' \in \mathcal{H}$, $d(h)$ is not a prefix of $d(h')$. Show that if $d$ is prefix-free then $\mathsf{VC}(\mathcal{H}) \leq \max_{h\in\mathcal{H}} |d(h)|$. [10 pts]

**2. (a)** Let $\mathcal{X}$ be a set, $\mathcal{H}$ be a finite class of functions $\mathcal{X} \to \{0,1\}$, $\mathcal{D}$ be a distribution over $\mathcal{X} \times \{0,1\}$, $A$ be a learning algorithm, $S \sim \mathcal{D}^m$, and let $h = A(S) \in \mathcal{H}$. Prove that for a uniform prior over $\mathcal{H}$, the PAC-Bayes theorem implies that for any $\delta \in (0,1)$, with probability at least $1 - \delta$,

$$L_{\mathcal{D}}(h) \leq L_S(h) + \sqrt{\frac{\ln(|\mathcal{H}|) + \ln(m/\delta)}{2(m-1)}}$$

with respect to the the 0-1 loss.

How does this bound compare quantitatively to the bound for finite hypothesis classes that we saw in the beginning of the course? [10 pts]

**(b)** Use the PAC-Bayes theorem to derive a bound similar to the MDL bound we saw in class (Theorem 7.7 in SSBD). How does this bound compare quantitatively to Theorem 7.7? [10 pts]

**3. (a)** Construct a class $\mathcal{H}$ of functions $\mathbb{N} \to \{0,1\}$ that is nonuniform PAC learnable but is not PAC learnable and prove that it satisfies these requirements. [10 pts]

**(b)** Let $\mathcal{X}$ be a set, let $\mathcal{H}$ be a class of functions $\mathcal{X} \to \{0,1\}$, and let $\{\mathcal{H}_n\}_{n\in\mathbb{N}}$ be a sequence of classes of functions $\mathcal{X} \to \{0,1\}$ such that $\mathcal{H} = \bigcup_{n\in\mathbb{N}} \mathcal{H}_n$. Prove that if $\mathcal{H}$ shatters a set $X \subseteq \mathcal{X}$ such that $|X| = \infty$ then there exists $k \in \mathbb{N}$ such that $\mathsf{VC}(\mathcal{H}_k) = \infty$. [30 pts]

*Hint: Let $\{X_n\}_{n\in\mathbb{N}}$ be disjoint subsets of $\mathcal{X}$ such that $|X_n| > \mathsf{VC}(\mathcal{H}_n)$ for all $n$. Show that for each $n \in \mathbb{N}$ there exists a function $f_n : X_n \to \{0,1\}$ that does not agree with any function in $\mathcal{H}_n$. Show that there exists a function $f : \mathcal{X} \to \{0,1\}$ that agrees with all the functions $f_n$. Argue that $f \in \mathcal{H} \setminus \bigcup_{n\in\mathbb{N}} \mathcal{H}_n$.*

**(c)** Construct a class $\mathcal{H}$ of functions $\mathbb{N} \to \{0,1\}$ that is not nonuniform PAC learnable and prove that it satisfies this requirement. [20 pts]