Assignment 2

Due: Monday, Nov 21, 12:15
Hand in at **beginning** of tutorial!

The solutions to this assignment will be discussed in the tutorial on Monday, Nov 21. If you want your assignment to be marked, it **has to be handed in at the beginning of the tutorial**. If you can not make it to the tutorial yourself, please arrange for one of your classmates to hand in your assignment.
A full score on this assignment are 50 marks. You can gain up to 10 bonus marks by answering the last question. To gain full marks, explain your answers carefully and prove your claims.

1. **Consistency**
   We say that a learner $\mathcal{A}$ is *consistent* with respect to a hypothesis class $\mathcal{H}$, if for all $\epsilon > 0$ and $\delta > 0$ and for every probability distribution P that is realizable by H, there is a sample size $m(\epsilon, \delta)$ such that for all $m > m(\epsilon, \delta)$

$$\mathbb{P}_{S \sim P^m}[\mathcal{L}(\mathcal{A}(S)) \ < \ \epsilon] \ \geq \ 1 - \delta$$

   In this exercise, you are asked to show that for *any* hypothesis class $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ any ERM algorithm for $\mathcal{H}$ is consistent with respect to $\mathcal{H}$ *if the domain is countable.*

   Let $\mathcal{X}$ be a countable domain and $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ be a hypothesis class and let P be a distribution over $\mathcal{X} \times \{0,1\}$ that is realizable by $\mathcal{H}$. For a labeled sample $S \in (\mathcal{X} \times \{0,1\})^m$, we let $S_{\mathcal{X}}$ denote its projection on the domain $\mathcal{X}$, that is $S_{\mathcal{X}} = \{x \in \mathcal{X} \ : \ (x,1) \in S \vee (x,0) \in S\}$.

   (a) Let $\{x_i \ : \ i \in \mathbb{N}\}$ be an enumeration of the elements in $\mathcal{X}$. Show that

$$\lim_{n \to \infty} P(\{x_n\}) \ = \ 0$$

   (b) Show that for any $\epsilon > 0$, there is an ϵ_P such that

$$P(\{x \in \mathcal{X} \ : \ P(\{x\}) < \epsilon_P\}) \ < \ \epsilon$$

   (c) Show that for every $\epsilon, \delta > 0$, there is an $m(\epsilon, \delta) \in \mathbb{N}$ such that for all $m > m(\epsilon, \delta)$

$$\mathbb{P}_{S \sim P^m}[P(\{x \in \mathcal{X} \ : \ x \notin S_{\mathcal{X}}) > \epsilon] \ < \ \delta$$

   (d) Show that for *any* hypothesis class $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ any ERM algorithm for $\mathcal{H}$ is consistent with respect to $\mathcal{H}$.

(5 + 5 + 5 + 5 marks)

2. **VC dimension of unions**
   Let H and H' be two hypothesis classes over some domain $\mathcal{X}$. Show that

$$\mathrm{VC}(H \cup H') \;\leq\; 2\max\{\mathrm{VC}(H), \mathrm{VC}(H')\} + 1$$

   Hint: Use Sauer's lemma.

(10 marks)

3. **Sauer's lemma**
   The proof of Sauer's lemma goes through showing that for any finite set $C = \{c_1, \ldots, c_m\} \subseteq \mathcal{X}$

$$|\mathcal{H}_C| \;\leq\; |\{B \subseteq C \;:\; \mathcal{H} \text{ shatters } B\}| \;\leq\; \sum_{i=0}^{d} \binom{m}{i}.$$

   Show that both of the above inequalities can be loose and can be tight. That is give examples of classes, where

   - $|\mathcal{H}_C| \;<\; |\{B \subseteq C \;:\; \mathcal{H} \text{ shatters } B\}|$
   - $|\mathcal{H}_C| \;=\; |\{B \subseteq C \;:\; \mathcal{H} \text{ shatters } B\}|$
   - $|\{B \subseteq C \;:\; \mathcal{H} \text{ shatters } B\}| \;<\; \sum_{i=0}^{d} \binom{m}{i}$
   - $|\{B \subseteq C \;:\; \mathcal{H} \text{ shatters } B\}| \;=\; \sum_{i=0}^{d} \binom{m}{i}$

   If you missed seeing the proof in the tutorial, you may look it up in the book "Understanding Machine Learning", page 49. Not needed for the exercise though.

(10 marks)

4. **VC-dimension**
   Consider the domain $\mathcal{X} = \mathbb{R}$ and the following hypothesis class over it:

$$\mathcal{H} = \{h_\theta \;:\; \theta \in \mathbb{R}\}, \quad h_\theta : x \mapsto \lceil \sin(\theta x) \rceil$$

   Prove that $\mathrm{VC}(\mathcal{H}) = \infty$.

(10 marks)

5. **VC-dimension of linear shifts**
   Consider the domain $\mathcal{X} = \mathbb{R}$. For a hypothesis class $\mathcal{H}$ over $\mathcal{X}$, we define its *shift class* $\mathcal{H}_s$ as

$$\mathcal{H}_s = \{h \in \{0,1\}^{\mathbb{R}} \;:\; h(x) = h'(x - a) \text{ for some } h' \in \mathcal{H} \text{ and } a \in \mathbb{R}\}$$

   Show that the gap between $\mathrm{VC}(\mathcal{H})$ and $\mathrm{VC}(\mathcal{H}_s)$ can be arbitrarily large (and even infinite).

(10 bonus marks)