Assignment 3

Due: Monday, Dec 19, 12:15
Hand in at **beginning** of tutorial!

The solutions to this assignment will be discussed in the tutorial on Monday, Dec 19. 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. **No-free-lunch**
   Prove the following (strengthened) version of the no-free-lunch theorem:

   (a) Let $\mathcal{X}$ be an infinite domain and let $\mathcal{A}$ be a learner. Let m be some sample size. Then, there exists a distribution P over $\mathcal{X} \times \{0, 1\}$ such that

   - there is a function $f : \mathcal{X} \to \{0, 1\}$ such that $L_P(f) = 0$
   - $\mathbb{E}_{S\sim P^m}[L_P(\mathcal{A}(S))] \geq 1/3$

   (b) Now argue how to strengthen the theorem even further. Show that, for every $a \in (0, 1/2)$, the second bullet point can be replaced by

   - $\mathbb{E}_{S\sim P^m}[L_P(\mathcal{A}(S))] \geq a$

   (10 + 5 marks)

2. **Perceptron**
   The Perceptron algorithm takes in a sequence of data points $S = ((x_1, y_1), \ldots, (x_m, y_m))$ and passes over this sequence data point by data point until none of the points in S is mislabeled.

   (a) Create a data sequence $S = ((x_1, y_1), \ldots, (x_m, y_m))$ (for example in $\mathbb{R}^2$) on which the Perceptron algorithm, after the first full pass over the data, will still make a mistake (during its second pass).

   (b) How many passes over the data will the algorithm make at most? Provide an upper bound.

   (10+5 marks)

3. **Uniform convergence**
Let $\mathcal{H} \subseteq \{0,1\}^{\mathcal{X}}$ be a hypothesis class. We say that $\mathcal{H}$ *has the uniform convergence property* if the following holds:

For all $\epsilon, \delta > 0$, there exists a $m(\epsilon, \delta) \in \mathbb{N}$ such that for all distributions P over $\mathcal{X} \times \{0,1\}$ and all $m \geq m(\epsilon, \delta)$, we have

$$\mathbb{P}_{S \sim P^m} \left[\max_{h \in \mathcal{H}} \{|L(h) - \hat{L}_m(h)|\} \leq \epsilon \right] \geq 1 - \delta$$

(a) Give 2 examples of classes that have the uniform convergence property. (You don't need to provide a full proof, just refer to the "right spot" in the lecture notes.)

(b) Show that, if classes $\mathcal{H}$ and $\mathcal{H}'$ have the uniform convergence property, then so does $\mathcal{H} \cap \mathcal{H}'$.

(c) Show that, if classes $\mathcal{H}$ and $\mathcal{H}'$ have the uniform convergence property, then so does $\mathcal{H} \cup \mathcal{H}'$.

(d) Show that, if a class $\mathcal{H}$ has the uniform convergence property, then so does the class $\mathcal{H}^c$ of complements of $\mathcal{H}$, defined as

$$\mathcal{H}^c := \{h \in \{0,1\}^{\mathcal{X}} \mid \text{there is a } h' \in \mathcal{H} \text{ such that } h(x) = 1 - h'(x) \text{ for all } x\}$$

(5+5+5+5 marks)

(e) [**Bonus**] Prove of refute: If a class $\mathcal{H}$ has the uniform convergence property, then so does its complement

$$\bar{\mathcal{H}} = \{0,1\}^{\mathcal{X}} \setminus \mathcal{H}$$

(10 marks)