Problem Sheet 3

Instructions: The problem sheets are designed to increase your understanding of the material taught in the lectures, as well as to prepare you for the final exam. You should attempt to solve the problems on your own after reading the lecture notes and other posted material, where applicable. Problems marked with an asterisk are optional. Once you have given sufficient thought to a problem, if you are stuck, you are encouraged to discuss with others in the course and with the lecturer during office hours. You are *not permitted* to search for solutions online.

1 VC Dimension of Linear Halfspaces in $\mathbb{R}^n$

We will show that the concept class of linear halfspaces in $\mathbb{R}^n$ has VC-dimension $n + 1$.

1. Give a set of $n + 1$ points in $\mathbb{R}^n$ that is shattered by the class of linear halfspaces.
2. We want to show that no set of $m = n + 2$ points in $\mathbb{R}^n$ can be shattered by the class of linear halfspaces. For this you can use what is called as Radon's theorem, described below.

3.* Prove Radon's theorem.

Given a set $S = \{x_1, \dots, x_m\} \subset \mathbb{R}^n$, the convex hull of S is the set,

$$\{z \in \mathbb{R}^n \mid \exists \lambda_1, \dots, \lambda_m \in [0, 1], \sum_{i=1}^{m} \lambda_i = 1, z = \sum_{i=1}^{m} \lambda_i x_i\}$$

Radon's Theorem: Let $m \geq n + 2$, then S must have two disjoint subsets S_1 and S_2 whose convex hulls intersect.

2 Properties of AdaBoost

Consider the AdaBoost algorithm as described in the lecture notes.

1. Show that the error of h_t with respect to the distribution D_{t+1} is exactly $1/2$.
2. What is the maximum possible value of $D_t(i)$ for some $1 \leq t \leq T$ and $1 \leq i \leq m$?
3. Fix some example, say i, let t_i be the first iteration such that $h_{t_i}(x_i) = y_i$. How large can t_i be?

3 Weak Learning CONJUNCTIONS and PARITIES

Consider the instance space $X_n = \{0, 1\}^n$. Consider the following hypothesis class:

$$H_n = \{0, 1, x_1, \bar{x}_1, x_2, \bar{x}_2, \dots, x_n, \bar{x}_n\}.$$

The hypothesis class contains $2n + 2$ functions. The functions "0" and "1" are constant and predict 0 and 1 on all instances in X_n. The function "x_i" evaluates to 1 on any $a \in \{0,1\}^n$ satisfying $a_i = 1$ and 0 otherwise. Likewise, the function "$\bar{x}_i$" evaluates to 1 on any $a \in \{0,1\}^n$ satisfying $a_i = 0$ and 0 otherwise. Thus a single bit of the input determines the value of these functions; for this reason these functions are sometimes referred to as *dictator* functions.

1. Show that the class CONJUNCTIONS is $\frac{1}{10n}$-weak learnable using H.

   Hint: The factor 10 is not particularly important, just a sufficiently large constant.

2. Let $\mathsf{CONJUNCTIONS}_k$ denote the class of conjunctions on at most k literals. Give an algorithm that PAC-learns $\mathsf{CONJUNCTIONS}_k$ and has sample complexity polynomial in k, $\log n$, $\frac{1}{\epsilon}$ and $\frac{1}{\delta}$. What would be the sample complexity if you had used the algorithm for learning CONJUNCTIONS discussed in the lectures?

   Hint: First show that the weak learning algorithm in the previous part can be modified to be a $\frac{1}{10k}$*-weak learner in this case.*

3. Show that there does not exist a weak learning algorithm for PARITIES using H.