CS/CNS/EE 253 - Advanced Topics in Machine Learning
Problem Set 2

Handed out:	4 Feb 2010
Due:	19 Feb 2010

---

1 VC Dimension

1. **Unions of intervals.** Consider data x that lies in the interval $[0, 1]$. Suppose the hypothesis space consists of indicator functions of unions of two intervals. More specifically, let the hypothesis space be parameterized by a, b, c, d satisfying $a < b$ and $c < d$ so that data points that fall in either interval $[a, b]$ or $[c, d]$ are classified as positive, and data that falls outside both intervals is classified as negative. What is the VC dimension of this hypothesis class?

2. **Decision Stumps in $\mathbb{R}^D$.** Decision stumps are a particularly simple family of binary classifiers for data $\mathbf{x}$ that lies in $\mathbb{R}^D$. Their classification rule has parameters q, i, α and takes the form $f(\mathbf{x}; i, q, \alpha) = q * \text{sign}(\mathbf{x}_i - \alpha)$. Decision stumps classify example $\mathbf{x}$ based only on the value of its i-th coordinate. α is a threshold value in $\mathbb{R}$ and q is either $+1$ or -1.

   (a) Consider n non-overlapping data points lying in $\mathbb{R}^D$. What is the maximum number of ways they can be classified using the decision stump family? That is, how many different binary labelings of the n points are there in the decision stump hypothesis space? Your result should be a function of n and D.

   (b) Show that the above result implies the following about the VC dimension of decision stumps:

$$VC_{ds} < 2(\log_2 D + 1) \tag{1}$$

2 Active Learning

The purpose of this question is to design an active learning strategy for the the "unions of intervals" hypothesis space from Problem 1.1. To simplify things, assume that $0 < a < b < c < d < 1$. Also assume that the parameters a, b, c, d are all separated by at least $\eta > 0$ and both intervals are at least length η such that, i.e., $a > \eta$, $b - a > \eta$, $c - b > \eta$, $d - c > \eta$, and $d < 1 - \eta$. Hereby, η is a constant, known to the algorithm. Suppose that the distribution over the inputs $P(x)$ is the uniform distribution over $[0, 1]$.

(a) Develop an active learning scheme that only requires $O(\log \frac{1}{\varepsilon} \log \frac{1}{\delta})$ labels to find a hypothesis with error at most ε with probability $1 - \delta$. Bound the number of labels that your algorithm requires as a function of ϵ, δ and η.

(b) Generalize this scheme to the hypothesis class consisting of hypotheses that are indicator functions of unions of k intervals, i.e., for each hypothesis h there exists $a_1, \ldots, a_k$, $b_1, \ldots, b_k$, $b_i - a_i > \eta$, $a_{i+1} - b_i > \eta$ and $a_1 > \eta$, $b_k < 1 - \eta$, such that h classifies inputs x positive if x is contained in one of the intervals $[a_i, b_i]$, negative otherwise.