Homework 6, Excercise 5.8

March 2, 2011

1. **(a)**

   The dimension $d = 1$.

   We will use adaboost to construct the majority vote over decision stumps. To prove that this will work we need to show that there is a constant $\gamma > 0$ such that for any distribution over the examples, there exists a decision stumps whose error is either smaller than $1/2 - \gamma$ or larger than $1/2 + \gamma$ (in which case we can use the inverse of the stump).

   Let the training set size be m. Then under any distribution D there is at least one instance a whose weight is at least $1/m$. Consider the two stump rules $f_1(x) = \mathbf{1}(x < a - \epsilon)$ and $f_2(x) = \mathbf{1}(x < a + \epsilon)$ where ϵ is small enough that only the instance in the interval $[a - \epsilon, a + \epsilon]$ is a. As the instance a is the only instance in the training set on which f_1 and f_2 differ, and as it's weight is at least $1/m$ the difference between the weighted training errors of the two rules satisfies $|\text{err}(f_1) - \text{err}(f_2)| > 1/m$. This implies that for at least one of $i = 1, 2$, $|\text{err}(f_i) - 1/2| > 1/(2m)$. We thus always have a stump with advantage $\gamma = 1/(2m)$.

   Requiring that $\epsilon = 1/m$ guarantees that we get a consistent rule using

   $$n = \frac{1}{2\gamma^2} \ln 1/\epsilon = 2m^2 \ln m$$

   Clearly a large overkill, but it works.

2. $d \geq 2$

   For $d = 2$ we will show that it is not possible to find a consistent classifier for the following set of examples (x_1, x_2, y) where x_1, x_2 are the coordinates of the point and y is the label $y \in \{-1, +1\}$

   $$(1, 1, +1), (-1, -1, +1), (-1, +1, -1), (+1, -1, -1)$$

   To prove that this is not possible we will use a statement that is a kind of an inverse to boosting. Suppose $\mathcal{H}$ is a set of base (or weak) classifiers mapping from X to $\{-1,+1\}$ and let $(x_1, y_1), \ldots, (x_n, y_n)$ be a training set where $y_i \in \{-1,+1\}$. Suppose there exists a weighted average of base classifiers $H(x) = \text{sign}(\sum_i \alpha_i h_i(x))$, $\alpha_i \geq 0$ that is consistent with the training set . Then for any distribution $\{p_1, \ldots, p_n\}$ over the training set thre exists a base classifier $h \in \mathcal{H}$ whose weighted error on the training set is smaller than $1/2$.

Using this claim it is easy to show that the training set described above cannot be represented using a weighted majority of stumps. Consider the uniform weighting over the four examples. It is clear that for this weighting neither a stump on x_1 not a stump on x_2 can have error smaller than $1/2$, which, using the claim, implies that there is no weighted majority of stumps that is consistent with this training set.

Proof of claim:
The fact that H is consistent with the training set implies that

$$\forall 1 \leq j \leq n, \;\; y_j \sum_i \alpha_i h_i(x_j) > 0$$

Summing this inequality over all the examples in the training set, weighted using $\{p_1, \ldots, p_n\}$ we get

$$\sum_{j=1}^{n} p_j y_j \sum_i \alpha_i h_i(x_j) > 0$$

which can be rewritten in the form

$$\sum_i \alpha_i \sum_{j=1}^{n} p_j y_j h_i(x_j) > 0$$

As all of the α_i are non-negative, there must exists at least one term in the external sum that is positive, i.e., there exists a value of i for which

$$\sum_{j=1}^{n} p_j y_j h_i(x_j) > 0$$

Which is in tern equivalent to

$$\sum_{j=1}^{n} p_j 2\mathbf{1}\left(h_i(x_j) = y_i\right) - 1 > 0$$

or in other words

$$\sum_{j=1}^{n} p_j \mathbf{1}\left(h_i(x_j) = y_i\right) > 1/2$$

As desired.