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 stump 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 $f_1(x) = 1(x < a - \epsilon)$ and $f_2(x) = 1(x < a + \epsilon)$ where $\epsilon$ 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 $H(x) = \text{sign} \left( \sum \alpha_i h_i(x) \right)$, $\alpha_i \geq 0$ that is consistent with the training set. Then for any distribution $\{p_1,\ldots,p_n\}$ over the training set there exists a base classifier $h \in H$ 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 $x_1$ nor 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

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

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

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

which can be rewritten in the form

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

As all of the $\alpha_i$ 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 turn equivalent to

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

or in other words

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

As desired.