Discussion 3: Boosting
Probabilistic Machine Learning,Fall 2016

1 Concepts

(a) In AdaBoost, if a weak learner gets example i wrong, then on the next round, the weight of example i will be: (higher, lower, or unchanged)?

(b) T/F: AdaBoost will eventually give zero training error regardless of the type of weak classifier it uses, provided enough iterations are performed. Do not assume the weak learning algorithm necessarily holds, simply assume we start with a dataset whose properties are not known.

2 Practice

1. Consider the following 2-class binary problem in Fig. 1. Using the update rules of the discrete AdaBoost, answer the following question

   (a) What are the first 2 decision stumps, and what are their corresponding thresholds? A decision stump is a variable with a single split ($x > c$, $x \leq c$).

   (b) Are these 2 stumps sufficient to obtain perfect classification?

2. Selecting the proper weak learning algorithm is often a very important step when tackling new classification problems. Fig. 2 present a number of such problems; for each dataset: a) Propose a set of weak classifiers that will excel in classifying the data b) Estimate the number of weak classifiers necessary to perform the task optimally

3. AdaBoost can be used in two ways. The first way is to combine a set of weak classifiers (such as stumps) where the set is determined before we run AdaBoost. The second way to run AdaBoost is as a meta algorithm on top of a weak learning algorithm like CART or C4.5. In that case, we can never actually enumerate all the weak classifiers. Look through all the steps of the AdaBoost algorithm and show that no step in the algorithm requires us to actually enumerate the set of weak classifiers. This means we never need to evaluate the matrix of margins $\mathbf{M}$. You will need to know that we do not actually need to obtain the best possible weak classifier (the argmax) in each round of AdaBoost in practice. It is sufficient to get a good weak classifier.

4. Assume the weak learning assumption holds. Asymptotically, as AdaBoost iterates over rounds, it is possible to determine what the values of $f(\mathbf{x}_i)$ will converge to?

$$
\begin{aligned}
C_{y=-1} &= \{(-3,-1),(-3,1),(3,-1),(3,1)\} \\
C_{y=1} &= \{(-1,-1),(-1,1),(1,-1),(1,1)\}
\end{aligned}
$$

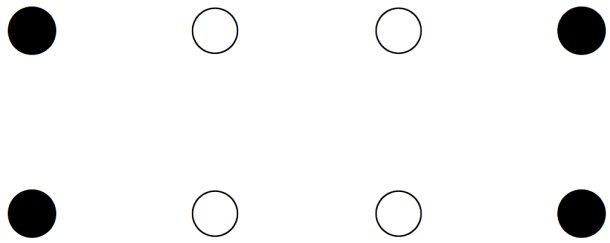

Figure 1: Toy data.

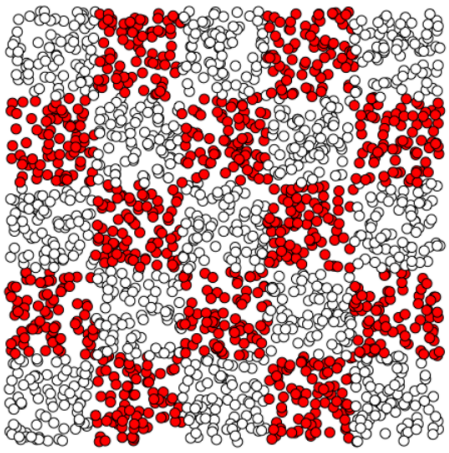

Figure 1: Checkerboard

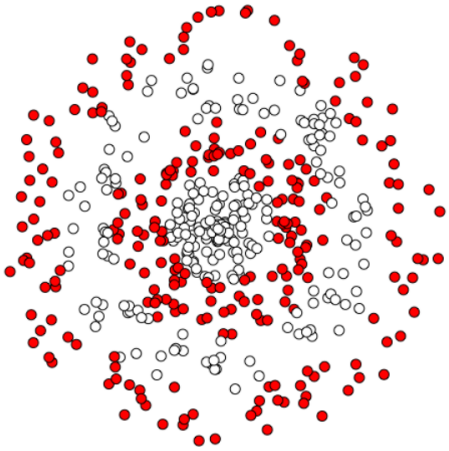

Figure 2: Concentric rings

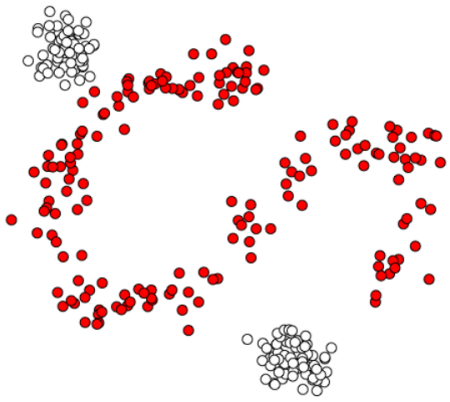

Figure 3: The serpent

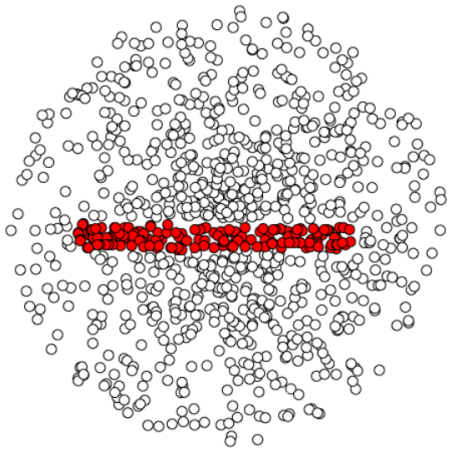

Figure 4: Stop sign

Figure 2: Toy data.