Machine Learning

CSCI 567 Spring 2021

Discussion: Boosting

Q1 Boosting Consider the AdaBoost algorithm shown below wherein the base algorithm is simply searching for a classifier with the smallest weighted error from a fixed classifier set $\mathcal{H}$.

Algorithm 1: Adaboost

Given: A training set $\{(\boldsymbol{x}_n, y_n \in \{+1, -1\})\}_{n=1}^N$, and a set of classifier $\mathcal{H}$, where each $h \in \mathcal{H}$ takes a feature vector as input and outputs $+1$ or -1.

Goal: Learn $H(\boldsymbol{x}) = \text{sgn}\left(\sum_{t=1}^T \beta_t h_t(\boldsymbol{x})\right)$, where $h_t \in \mathcal{H}$ and $\beta_t \in \mathbb{R}$.

Initialization: $D_1(n) = \frac{1}{N}$, $\forall n \in [N]$.

for $t = 1, 2, \cdots, T$ **do**

  Find $h_t = \arg\min_{h \in \mathcal{H}} \sum_{n: y_n \neq h(\boldsymbol{x}_n)} D_t(n)$.

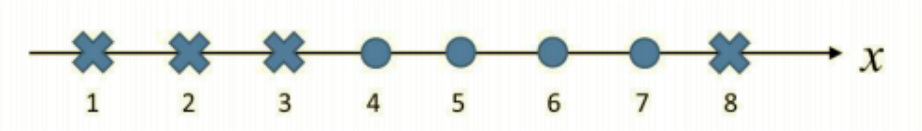
  Compute

$$\epsilon_t = \sum_{n: y_n \neq h_t(\boldsymbol{x}_n)} D_t(n) \qquad \text{and} \qquad \beta_t = \frac{1}{2} \ln \frac{1 - \epsilon_t}{\epsilon_t}.$$

  Compute for each $n \in [N]$,

$$D_{t+1}(n) \propto D_t(n) \exp(-\beta_t y_n h_t(\boldsymbol{x}_n)).$$

Imagine running AdaBoost with a 1-d training set of 8 examples, where circles mean $y = +1$ and crosses mean $y = -1$. The number under each example is its x coordinate.



The base classifier set $\mathcal{H}$ consists of all decision stumps, where each of them is parameterized by a pair $(s, b) \in \{+1, -1\} \times \mathbb{R}$ such that

$$h_{(s,b)}(x) = \begin{cases} s & \text{if } x > b, \\ -s & \text{otherwise} \end{cases}$$

(a) Which of the following is a possible parameter for h_1?

(A) (s, b) = (+1, 3.5)

(B) (s, b) = (-1, 3.5)

(C) (s, b) = (+1, 7.5)

(D) (s, b) = (-1, 7.5)

(b) Suppose we run AdaBoost for two rounds and observe that β_1 and β_2 are both positive but not equal. Is it possible that the final classifier H (line 2) after these two rounds has zero training error? Why or why not?