Relative Mistake Bound for Weighted-Majority

Geoffrey Ulman
Homework 10
CSI873

November 2011

1 Theorem

Relative mistake bound for Weighted-Majority. Let D be any sequence of training examples, let A be any set of n prediction algorithms, and let k be the minimum number of mistakes made by any algorithm in A for the training sequence D. Then the number of mistakes over D made by the Weighted-Majority algorithm using $0 < \beta < 1$ is given by Equation 1.

$$\frac{-k \log_2(\beta) + \log_2(n)}{1 - \log_2(1+\beta)} \tag{1}$$

2 Proof

The weight for a given prediction algorithm a_j after making k mistakes is given by β^k.

Let W be the sum of the weights for all n algorithms. Initially, $W = n$ since the weights w_j of each prediction alagorithm are initialized to 1 (Equation 2).

$$\sum_{i=1}^{n} w_j = W \tag{2}$$

When the Weighted-Majority learner makes a mistake, the total weight W will be reduced most when all n prediction algorithms make mistakes. In this case each has their weight reduced by β. The total weight W will be reduced the least in the case where prediction algorithms accounting exactly half of the total weight make mistakes. In this case, the total weight will be reduced by $\frac{1}{2} + \frac{\beta}{2}$ or $\frac{1+\beta}{2}$.

Now let M denote the total number of mistakes made by the Weighted-Majority learner. From above, the total weight after M mistakes cannot be greater than $n\left(\frac{1+\beta}{2}\right)^M$. Further, this quantity must be greater than the weight w_j of any individual learner, giving Equation 3.

$$\beta^k \leq n\left(\frac{1+\beta}{2}\right)^M \tag{3}$$

Taking the base-2 log of both sides results in Equaion 4.

$$k\log_2(\beta) \leq \log_2(n) + M\log_2\left(\frac{1+\beta}{2}\right) \tag{4}$$

Move the $\log_2(n)$ term to the other side of the inequality and divide by $\log_2\left(\frac{1+\beta}{2}\right)$ to reach Equation 5. Note that the divisor is always negative because $0 < \beta < 1$, so the inequality is flipped.

$$\frac{k\log_2(\beta) - \log_2(n)}{\log_2\left(\frac{1+\beta}{2}\right)} \geq M \tag{5}$$

Multiplying the numerator and denominator by -1 yields Equation 6

$$\frac{-k\log_2(\beta) + \log_2(n)}{\log_2\left(\frac{2}{1+\beta}\right)} \geq M \tag{6}$$

Simplifying the denominator yields Equation 7 which is the expression given by the theorem to be proven.

$$\frac{-k\log_2(\beta) + \log_2(n)}{1 - \log_2(1+\beta)} \geq M \tag{7}$$

References

[1] Tom M. Mitchell, *Machine Learning*, WCB McGraw-Hill, Boston, 1997.