Winnow (algorithm)

From Wikipedia, the free encyclopedia
(Redirected from Winnow algorithm)

The **winnow algorithm**[1] is a technique from machine learning for learning a linear classifier from labeled examples. It is very similar to the perceptron algorithm. However, the perceptron algorithm uses an additive weight-update scheme, while Winnow uses a multiplicative scheme that allows it to perform much better when many dimensions are irrelevant (hence its name). It is a simple algorithm that scales well to high-dimensional data. During training, Winnow is shown a sequence of positive and negative examples. From these it learns a decision hyperplane that can then be used to label novel examples as positive or negative. The algorithm can also be used in the online learning setting, where the learning and the classification phase are not clearly separated.

Algorithm [edit]

The basic algorithm, Winnow1, is as follows. The instance space is $X = \{0,1\}^n$, that is, each instance is described as a set of Boolean-valued features. The algorithm maintains non-negative weights w_i for $i \in \{1...n\}$, which are initially set to 1, one weight for each feature. When the learner is given an example $(x_1, ...x_n)$, it applies the typical prediction rule for linear classifiers:

- **If** $\sum_{i=1}^{n} w_i x_i > \Theta$, **then** predict 1
- **Otherwise** predict 0

Here Θ is a real number that is called the *threshold*. Together with the weights, the threshold defines a dividing hyperplane in the instance space. Good bounds are obtained if $\Theta = n/2$ (see below).

For each example with which it is presented, the learner applies the following update rule:

- If an example is correctly classified, do nothing.
- If an example is predicted to be 1 but the correct result was 0, all of the weights implicated in the mistake are set to 0 (demotion step).
- If an example is predicted to be 0 but the correct result was 1, all of the weights implicated in the mistake are multiplied by α (promotion step).

Here, "implicated" means weights on features of the instance that have value 1. A typical value for α is 2.

There are many variations to this basic approach. *Winnow2*[1] is similar except that in the demotion step the weights are divided by α instead of being set to 0. *Balanced Winnow* maintains two sets of weights, and thus two hyperplanes. This can then be generalized for multi-label classification.

Mistake bounds [edit]

In certain circumstances, it can be shown that the number of mistakes Winnow makes as it learns has an upper bound that is independent of the number of instances with which it is presented. If the Winnow1 algorithm uses $\alpha > 1$ and $\Theta \geq 1/\alpha$ on a target function that is a k-literal monotone disjunction given by $f(x_1, ...x_n) = x_{i_1} \cup ... \cup x_{i_k}$, then for any sequence of instances the total number of mistakes is bounded by: $\alpha k (\log_\alpha \Theta + 1) + \frac{n}{\Theta}$.[2]

References [edit]

1. ^ a b Nick Littlestone (1988). "Learning Quickly When Irrelevant Attributes Abound: A New Linear-threshold Algorithm", *Machine Learning 285–318(2)*.
2. ^ Nick Littlestone (1989). "Mistake bounds and logarithmic linear-threshold learning algorithms". Technical report UCSC-CRL-89-11, University of California, Santa Cruz.

Categories: Classification algorithms