Boosting:

Or How to Make a Silk Purse Out of a Pig's Ear

Harrison B. Prosper

Florida State University

Single Top Group

20 April 2005

Outline

- Introduction
- Adaptive Boosting
- Summary

Introduction

- Boosting
  - This is a general method for improving the performance of *any* weak classifier.
  - A weak classifier performs only slightly better than random guessing!
  - In principle, boosting can be applied to
    - NN
    - KNN
    - Trees
    - etc.

Boosting – General Algorithm

$T_1 = [(x_1, y_1, w_1), \ldots (x_N, y_N, w_N)] = \text{initialize}()$

for k in 1,…K

$f_k(x) = \text{train}(T_k)$

$\alpha_k = \text{minimize}\ (\text{ErrorFunction}_k)$

$T_{k+1} = \text{modify}(T_k)$

$$F(x) = \sum_{k=1}^{K} \alpha_k f_k(x)$$

Adaptive Boosting – Algorithm

$T_1 = [(x_1, y_1, w_1), \ldots (x_N, y_N, w_N)]$ = initialize()

for k in 1,…K

$f_k(x)$ = train(T_k)

compute $e_k = \sum_{i:\, y\, f\, <\, 0} w_i$

compute $\alpha_t = \ln [(1-e_k)/e_k]$

update $w_{k+1,\, i} = w_{k,\, i} \exp[-\alpha_k\, y_k f_k(x_i)/2]$ / Z_k

$$F(x) = \sum_{k=1}^{K} \alpha_k f_k(x)$$

Adaptive Boosting

- Training
  - Training data $(x_1, y_1), (x_2, y_2), \ldots$
  - y is -1 for background
  - y is $+1$ for signal.

  - Train a classifier $f(x)$ that assigns -1 or $+1$ to x.

Adaptive Boosting – II

- Choosing α_t
  - Consider the product $u_i = y_i f(x_i)$ for some event x_i with class label y_i and weight w_i
    - If $f(x) = y$ then $u = +1$
    - If $f(x) \neq y$ then $u = -1$
  - For a given classifier $f(x)$, this suggests choosing α such that

    is minimized $$Z_k = \sum_{i=1}^{N} w_i \exp(-\alpha_k u_i / 2)$$

Adaptive Boosting – III

- Re-write Z in terms of error rate *err*:

$$Z = \sum_{i=1}^{N} w_i \exp(-\alpha u_i) = \exp(-\alpha) \sum_{i:u_i=+1} w_i + \exp(\alpha) \sum_{i:u_i=-1} w_i$$

$$= \exp(-\alpha)(1 - err) + \exp(\alpha)\, err$$

- Minimize:

$$\frac{dZ}{d\alpha} = -\exp(-\alpha)(1 - err) + \exp(\alpha)\, err = 0$$

- And find:

$$\alpha = \ln[(1 - err)/err]$$

Adaptive Boosting – IV

- Normalization Z

$$Z_k = 2\sqrt{\varepsilon_k(1-\varepsilon_k)}$$

- Training error is bounded by

$$\varepsilon \le \prod 2\sqrt{\varepsilon_k(1-\varepsilon_k)} < \exp(-2K\gamma^2)$$

if $0.5 - e_k > \gamma > 0$

γ is referred to as the weak edge. So, if one runs the algorithm forever, the training error $\rightarrow$ zero

Summary

- Boosting
  - This is simply another method to combine many classifiers to make one that works better than any individual.
  - In principle, it can be applied to any method.
  - But, if a method is already very powerful, boosting won't help much!