log & inv

Suppose hypothesis $h_t$ has error $\epsilon_t$. For simple boost, I tried two ways to assign weight to $h_t$:

- log: $1/\log(1.01 + \epsilon_t)$
- inv: $1/\epsilon_t$

They didn’t show much difference.

Possible causes

- $\epsilon_t$ is small ($0.015 \sim 0.045$), thus
  \[ \log(1.01 + \epsilon_t) \approx 0.01 + \epsilon_t \]
- boosting is insensitive to small weight changes (?)
Towards the final (best) hypothesis

- Use 12000 samples for training and 3997 for testing
- Select the one with smallest testing error out of no more than 3 candidates

Generalization

Let the out-of-sample error be \( \Pi \). The size of the never-touched test set is \( T \).

The classification error \( e_i \in \{0, 1\} \) for test sample \( i \) is an unbiased estimate of \( \Pi \). Thus the testing error

\[
\nu = \frac{1}{T} \sum_{i=1}^{T} e_i
\]

is also an unbiased estimate of \( \Pi \).

For big \( T \), \( \nu \) can be regarded as a Gaussian with mean \( \Pi \) and standard deviation

\[
\sigma = \sqrt{\frac{\pi(1-\pi)}{T}} \approx \sqrt{\frac{\nu(1-\nu)}{T}}.
\]

The 95% confidence interval of \( \Pi \) is

\[
[\nu - 1.96\sigma, \nu + 1.96\sigma]
\]

since

\[
\int_{1.96}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \approx 2.5%.
\]
One candidate

For AdaBoost.M2 with stochastic gradient descent, I got $\nu = 2.752\%$. Then with high probability $\Pi$ is within $[2.24\%, 3.26\%]$