AdaBoost and Exponential Loss

- so AdaBoost is greedy procedure for minimizing exponential loss

$$\prod_t Z_t = \frac{1}{m} \sum_i \exp(-y_i F(x_i))$$

where

$$F(x) = \sum_t \alpha_t h_t(x)$$

- why exponential loss?
  - intuitively, strongly favors $F(x_i)$ to have same sign as y_i
  - upper bound on training error
    - smooth and convex (but very loose)
- how does AdaBoost minimize it?