6.034 Exam 4 Cheat Sheet

Adaboost

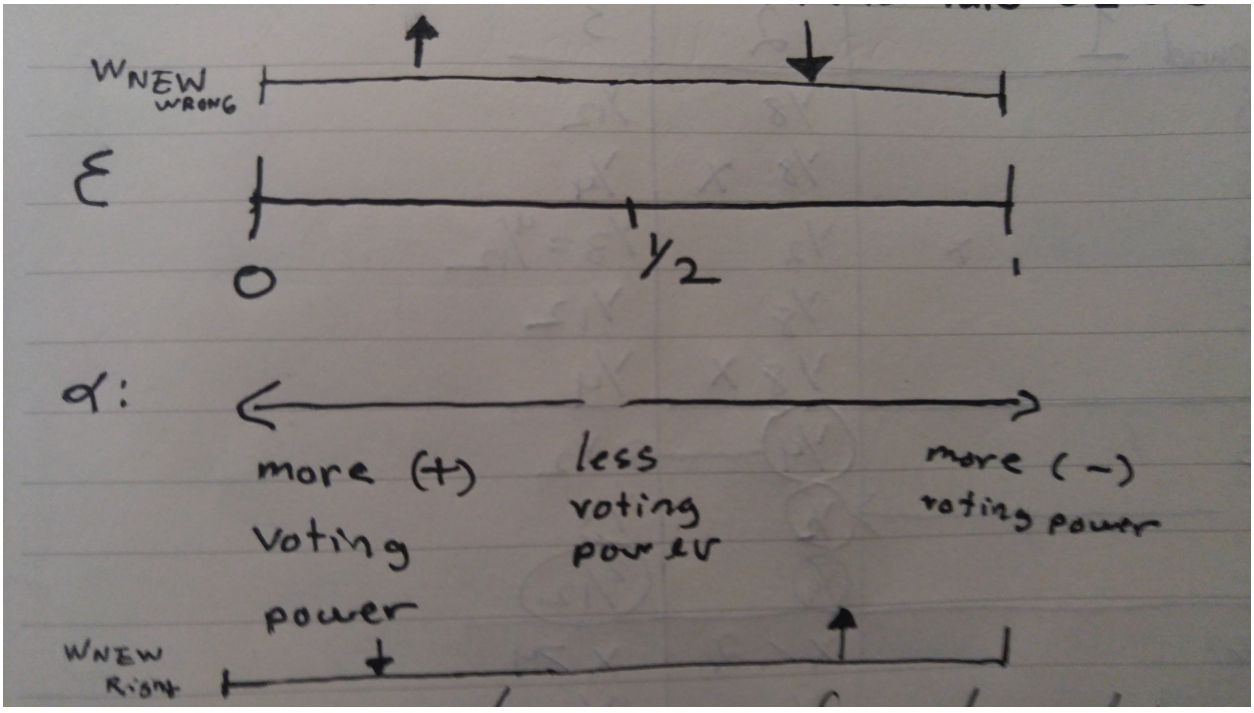
Overview

Take imperfect (weak) classifiers h_i to create a strong overall classifier:

$$H(\vec{x}) = \text{SIGN}(\underbrace{\alpha_1}_{\text{voting power}} \cdot h_1(\vec{x}) + \alpha_2 \cdot h_2(\vec{x}) + \ldots)$$

Each weak classifier h_i has an error rate, ϵ, where $0 \leq \epsilon \leq 1$

Error rate $\epsilon = \sum w_i$



Note: after selecting a weak classifier h_i, the points that were incorrectly classified by h_i have their w_i increase, while the points that were correctly classified by h_i have their w_i decrease

We build $H(\vec{x})$ in a series of rounds, each round:

- pick the best* h_i
- calculate voting power α_i
- append h_i to overall classifier, H

* best can mean two different things:

- smallest ϵ value
- ϵ value furthest away from $\frac{1}{2}$

Steps

0. Initialize $w_i = \frac{1}{N}$ (assign equal weight to each training points)
1. Find best* h w/ lowest error rate
2. Calculate voting power: $\alpha = \frac{1}{2} \ln \frac{1-\epsilon}{\epsilon}$
3. Check to see if we're finished:
   - H is good enough (perfectly classifies all of the training data)
   - no good weak classifiers left (i.e. best h_i has $\epsilon_i = \frac{1}{2}$ - no better than a coin-flip)
   - have performed enough rounds of Adaboost
4. Update Weights:

$$w_{\text{NEW}} = \begin{cases} \frac{1}{2} \times \frac{1}{1-\epsilon} \times w_{\text{OLD}} & \text{if } w_{\text{OLD}} \text{ was correctly classified by newly selected } h_i \\ \frac{1}{2} \times \frac{1}{\epsilon} \times w_{\text{OLD}} & \text{if } w_{\text{OLD}} \text{ was incorrectly classified by newly selected } h_i \end{cases}$$

(Somewhat) Random Helpful Facts

- If you have 3+ weak classifiers that make non-overlapping errors, you can make a perfect classifier
  - can't make a perfect classifier with less than 3
- Only after Round 1: $\sum_{\text{right}} w_i = \sum_{\text{wrong}} w_i = \frac{1}{2}$
- At any point: $0 < w_i \leq \frac{1}{2}$
  - notably: w_i can never be 0
- At any point: $\sum w_i = 1$
- w_i **must** change every round, since we aren't selecting classifiers with $\epsilon_i = \frac{1}{2}$
- After selecting h_i, on the following round: $\mathcal{E}(h_i) = \frac{1}{2}$
- Say we have two classifiers: h_1 and h_2, and our criteria for selecting the best classifier on any given round is choosing the one with the lowest overall ϵ_i. If h_1 misclassifies points A, B and h_2 misclassifies A, B, C, then h_2 will **never** be selected. h_2 has the superset of h_1's errors.
  - If the "best" criteria is the furthest ϵ_i from $\frac{1}{2}$ instead, then h_2 could be chosen.
- Adaboost tends **not** to overfit