CSE 446
Ensembles
Administrative

• Quiz sections
  – No quiz section this week
  – Will have additional quiz section next week (announced soon)
    • Midterm problems & answers
    • Differentiation (come to this if you found question 2 on the exam or backprop on homework 2 difficult!)

• Midterm grading in progress...
Boosting

[Schapire, 1989]

• **Idea:** if we give each weak learner a difference piece of the dataset, we get a really good complex classifier from letting them vote

• **Learners must be different** (how was this achieved in Bagging?)

• **Learners must be better than random** (not too weak)

• **Approach:** given a weak learner, run it multiple times on (rewighted) training data, then let learned classifiers vote

• **On each iteration** $t$:
  
  – weight each training example by how incorrectly it was classified
  
  – Learn a hypothesis – $h_t$
  
  – A strength for this hypothesis – $\alpha_t$

• **Final classifier:**

$$h(x) = \text{sign} \left( \sum_i \alpha_i h_i(x) \right)$$

• **Practically useful**

• **Theoretically interesting**
time = 0
blue/red = class
size of dot = weight
weak learner = Decision stub: horizontal or vertical
time = 1

this hypothesis has 15% error and so does this ensemble, since the ensemble contains just this one hypothesis

First, generate a data-set by clicking on the left and right buttons in the main window of the applet. Then, press "split" to split the data into training and test sets.
time = 2

First, generate a data-set by clicking on the left and right buttons in the main window of the applet. Then, press "split" to split the data into training and test sets.
time = 3
time = 13

First, generate a data-set by clicking on the left and right buttons in the main window of the applet. Then, press "split" to split the data into training and test sets.

Applet adaboost started
First, generate a data-set by clicking on the left and right buttons in the main window of the applet. Then, press "split" to split the data into training and test sets.
time = 300

overfitting

First, generate a data-set by clicking on the left and right buttons in the main window of the applet. Then, press "split" to split the data into training and test sets.
Boosting results – Digit recognition

[Schapire, 1989]

- **Boosting:**
  - Seems to be robust to overfitting
  - Test error can decrease even after training error is zero?

![Graph showing error vs. number of rounds with labels for training and test error](image-url)
Boosting generalization error bound

\[ \text{error}_{\text{true}}(H) \leq \text{error}_{\text{train}}(H) + \tilde{O}\left(\sqrt{\frac{Td}{m}}\right) \]

Constants:

- **\( T \)**: number of boosting rounds
  - Higher \( T \) \( \rightarrow \) Looser bound
- **\( d \)**: measures complexity of classifiers
  - Higher \( d \) \( \rightarrow \) bigger hypothesis space \( \rightarrow \) looser bound
- **\( m \)**: number of training examples
  - More data \( \rightarrow \) tighter bound

[Freund & Schapire, 1996]
Boosting generalization error bound

\[ error_{true}(H) \leq error_{train}(H) + \tilde{O}\left(\sqrt{\frac{Td}{m}}\right) \]

Constants:

- **Theory does not match practice:**
  - Robust to overfitting
  - Test set error decreases even after training error is zero
- **Need better analysis tools**
  - we’ll come back to this later in the quarter

- more data \(\Rightarrow\) tighter bound
Boosting: Experimental Results

Comparison of C4.5, Boosting C4.5, Boosting decision stumps (depth 1 trees), 27 benchmark datasets

[Freund & Schapire, 1996]